Filosofía contemporánea

La naturaleza de la experiencia

Volumen I
Sensaciones

Maite Ezcurdia y Olbeth Hansberg
(compiladoras)

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO ✣ INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

Serie antologías

¿Qué es lo que distingue una sensación de otra? ¿Un dolor de un cosquilleo, por ejemplo? Las sensaciones constituyen el mayor obstáculo para dar una explicación satisfactoria de la mente en términos no mentales. Éstas se distinguen por la manera como se sienten, esto es, por sus *qualia* o su carácter fenoménico, que se resiste a ser explicado, y que hace patente el hiato entre lo físico y lo mental. Pero, ¿en qué consiste este aspecto de las sensaciones? ¿Lograremos alguna vez dar cuenta de él adecuadamente? Éstas son dos de las preguntas fundamentales que se discuten en esta antología.

Los artículos de este volumen exploran diferentes maneras de dar cuenta de los *qualia* y de las dificultades que surgen al intentarlo. En particular, se sitúan en el debate actual sobre los problemas que los *qualia* presentan a las teorías fisicalistas, funcionalistas e intencionalistas de la mente. *La naturaleza de la experiencia* reúne doce artículos: el primero, de las editoras, ofrece un panorama general de la discusión contemporánea en torno a las sensaciones y su carácter fenoménico; los demás son traducciones de textos centrales en esta discusión, como los ya clásicos de Nagel ("¿Cómo es ser un murciélago?") y de Jackson ("*Qualia* epifenoménicos" y "Lo que María no sabía"), y trabajos más recientes de Harman, Block y Tye. Se incluyen, además, artículos de Dennett, Block y Fodor, McGinn, Lewis y Shoemaker.

ISBN 970-32-0545-1

Maite Ezcurdia es investigadora del Instituto de Investigaciones Filosóficas de la Universidad Nacional Autónoma de México. Realizó sus estudios de doctorado en la Universidad de Londres, y ha publicado diversos artículos en inglés y en español, principalmente, en las áreas de filosofía del lenguaje y filosofía de la mente. Actualmente prepara dos antologías: *Orayen: de la forma lógica al significado* y (con Robert Stainton y Chris Viger) *Language, Mind and World*.

La doctora en filosofía Olbeth Hansberg es investigadora del Instituto de Investigaciones Filosóficas y profesora de la Facultad de Filosofía y Letras de la UNAM. Ha publicado artículos sobre diversos temas de filosofía de la mente y del lenguaje, especialmente sobre las emociones. Es autora del libro *La diversidad de las emociones* (FCE, 1996), y compiladora, junto con Mark Platts, de *Responsabilidad y libertad* (UNAM/FCE, 2002). Desde 2000 es Coordinadora de Humanidades de la UNAM.

LA NATURALEZA DE LA EXPERIENCIA

VOLUMEN I: SENSACIONES

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS
Directora: DRA. PAULETTE DIETERLEN
Secretario Académico: DR. RICARDO SALLES

Colección: FILOSOFÍA CONTEMPORÁNEA
Serie: ANTOLOGÍAS

LA NATURALEZA DE LA EXPERIENCIA

VOLUMEN I: SENSACIONES

Compilación, introducción
y revisión técnica de las traducciones
MAITE EZCURDIA
y
OLBETH HANSBERG

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MÉXICO 2003

Agradecemos a los proyectos CONACYT 3128P-H "Conciencia y estados mentales", y CONACYT 27748-H "Los límites de la conciencia", el apoyo financiero para la traducción y elaboración de esta antología. Asimismo, agradecemos a los participantes en dichos proyectos las discusiones sobre el tema de sensaciones y *qualia*, a partir de las cuales se llevó a cabo esta selección.

DR © 2003, Universidad Nacional Autónoma de México

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS
Circuito Mtro. Mario de la Cueva s/n,
Ciudad Universitaria, Coyoacán, 04510, México, D.F.
Tels.: 56 22 74 37 y 56 22 75 04; fax 56 65 49 91
correo electrónico: libros@filosoficas.unam.mx
página web: http://www.filosoficas.unam.mx

Impreso y hecho en México
ISBN 970-32-0546-1

INTRODUCCIÓN

SENSACIONES, CONCIENCIA FENOMÉNICA Y LOS LÍMITES DE LA EXPLICACIÓN

MAITE EZCURDIA Y OLBETH HANSBERG

Cuando nos enfrentamos a un fenómeno, sea cual fuere, intentamos entenderlo y entender su lugar frente a los demás fenómenos, especialmente aquellos que ya comprendemos. Tal es el caso de los fenómenos físicos (por ejemplo, la erosión de las orillas de un río, la solubilidad de la sal, el funcionamiento del cerebro humano), así como de los fenómenos mentales. Entre estos últimos, buscamos explicaciones de estados mentales como creencias, deseos, emociones y de lo que podemos llamar "experiencia".

"Experiencia" es un término que se usa para denotar varios tipos de estados, sucesos o capacidades. A veces decimos que alguien tiene experiencia en arreglar autos, o que tiene gran experiencia en política. En estos casos estamos denotando capacidades o un conocimiento práctico que un sujeto ha adquirido. En otras ocasiones, se puede usar el término para denotar algo que le pasó a uno, como cuando decimos que un suceso constituyó una experiencia inolvidable. Ninguna de esas acepciones nos interesa aquí. Nuestro interés en la experiencia reside en aquellas acepciones que denotan sucesos o estados mentales. Se incluyen aquí los diversos estados perceptuales como ver la silla, oír una voz, y las sensaciones como dolores, ardores y cosquilleos, entre otros.

Al buscar explicaciones de la naturaleza de la experiencia tratamos de dar cuenta de los rasgos distintivos de la experiencia y de sus relaciones con otros estados o sucesos físicos y mentales. Intentamos encontrar respuestas a preguntas como ¿qué distingue a una experiencia perceptual de otros estados mentales, como creencias, deseos, y estados de ánimo (por mencionar algunos)?, ¿qué hace que esa experiencia constituya la percepción de un estado de cosas en el mundo? y ¿qué relación existe entre una experiencia perceptual y otros estados mentales y acciones?

Las experiencias perceptuales, por ejemplo, típicamente generan creencias en los sujetos acerca de sus entornos, las cuales, cuando son verídicas, sirven para guiar sus acciones en esos entornos. Las experiencias perceptuales además de poder ser verídicas o correctas, exhiben otras características de las que carecen otros estados mentales. En particular, las experiencias visuales se distinguen de otros estados mentales, incluyendo otros tipos de experiencias, por las sensaciones que las acompañan, por sensaciones de colores que les son constitutivas. Así pues, si queremos entender qué es una experiencia visual necesitamos caracterizaciones y explicaciones de sus relaciones con otros estados mentales, de cómo y en qué sentido puede ser verídica o correcta, y de qué la distingue.

La meta principal de esta colección sobre la naturaleza de la experiencia es ofrecer al lector dos conjuntos de artículos contemporáneos (casi todos traducidos por primera vez al español) que han tenido gran influencia en la discusión actual sobre los problemas de la experiencia perceptual. En el segundo volumen, nos abocaremos a la investigación de lo que hace que una percepción sea verídica o correcta (y qué que no lo sea) y de las relaciones (causales y justificatorias) que tiene con otros estados mentales. En este primer volumen nos ocupamos de uno de los problemas más difíciles con los que se ha enfrentado la filosofía, y que surge de algo que al parecer permea toda la experiencia, a saber, las sensaciones.[1] El problema surge a partir de una característica que comparten todas las sensaciones y que

[1] Decimos "al parecer" porque el caso de visión ciega podría considerarse como un caso en el que hay cierto tipo de percepción sin sensaciones: un sujeto reporta no poder ver en un área de su campo visual y, no obstante, parece procesar información visual de esa área. Dejamos para el segundo volumen la

se ha denominado "la conciencia fenoménica",[2] "*qualia*", "el aspecto subjetivo de la experiencia", "el contenido cualitativo" o "el contenido fenoménico" de la experiencia, entre otros.

Ejemplos de sensaciones son las sensaciones de los colores, los dolores, los ardores, cosquilleos u hormigueos en la piel. Algunas de las sensaciones constituyen ellas mismas percepciones. Esto sucede, por ejemplo, cuando al ver la pared frente a mí experimento la sensación de blanco. La sensación que en este caso experimento constituye mi percepción de la pared blanca.[3] Pero a primera vista no todas las sensaciones son percepciones. El cosquilleo u hormigueo que siento sobre la piel no parece serlo, ni tampoco el dolor de cabeza. Lo que todas las sensaciones comparten (sean o no percepciones) es el hecho de que cada una tiene cierta fenomenología o *quale*. La conciencia fenoménica (el *quale*) asociada con un dolor es *la manera como se experimenta o se siente* dicho dolor. El dolor que siento cuando se me ha roto el apéndice es distinto del dolor que siento después de un golpe en la región del apéndice. La diferencia reside en cómo experimento o siento dicho dolor, la diferencia está en su fenomenología o en los distintos *qualia* asociados. De la misma manera, mi sensación de un tono de rojo, digamos, el rojo 13, difiere de mi sensación de otro tono, el rojo 14, no sólo por los tonos mismos, sino por la manera en que experimento dichos tonos, por su fenomenología.

En esta introducción explicamos la selección de artículos en este volumen y el orden de su presentación, lo cual esperamos

pregunta de si todas las percepciones están acompañadas por sensaciones o están constituidas por ellas.

[2] Es importante distinguir la conciencia fenoménica de otras nociones de conciencia, en particular, de la noción que acompaña a frases como "soy consciente de mis creencias", la cual ha llegado a caracterizarse como conciencia de segundo orden (véase Rosenthal 1986). Dos excelentes artículos que presentan caracterizaciones de diferentes nociones de conciencia y que además presentan explicaciones de cómo éstas interactúan con la conciencia fenoménica son Block 1995 y Burge 1997. Una excelente colección en inglés sobre diferentes temas y problemas que presenta la conciencia es Block, Flanagan y Güzeldere 1997. En lo que sigue al hablar de conciencia nos estaremos refiriendo a la conciencia fenoménica, a menos que se señale lo contrario.

[3] Dejaremos para el segundo volumen la discusión sobre las condiciones necesarias y suficientes para que una sensación constituya una percepción.

sirva como guía para su lectura y discusión, y para presentar nuestra propia visión sobre el estado actual del debate.

1. *Subjetividad y los límites explicativos*

De manera general y no exhaustiva, podemos agrupar las explicaciones en dos grandes clases: las explicaciones *causales* y las explicaciones *constitutivas* o *composicionales*. Las explicaciones causales nos permiten entender cómo se generó un fenómeno a partir de otro(s), mientras que las explicaciones constitutivas se centran en la pregunta sobre cuáles elementos constituyen un fenómeno. Estas explicaciones no tienen por qué ser excluyentes. Por ejemplo, al ofrecer una explicación de cómo se obtiene agua dulce a partir de la inmersión de un cubo de azúcar en agua pura, se ofrece una explicación causal que puede apelar a componentes del azúcar, esto es, a su composición química, y a la composición química del agua para explicar la solubilidad del azúcar en el agua. Así, dos explicaciones composicionales o constitutivas pueden servir para explicar causalmente cómo se obtuvo agua dulce.

Al intentar comprender la conciencia fenoménica o el aspecto cualitativo de las sensaciones, tendremos que ofrecer explicaciones que sean causales y constitutivas. Las explicaciones causales nos dirán cómo las sensaciones surgen a partir de ciertos sucesos físicos o mentales y cuáles son sus conexiones con otros sucesos. Sin embargo, si estas explicaciones han de ser iluminadoras nos deberán decir no sólo qué causa qué sino cómo lo hace. Para ello, como en el caso del azúcar disuelta en el agua, será necesario ofrecer explicaciones constitutivas de un estado de experiencia —del estado en el que estoy al experimentar una sensación—, las cuales nos permitan decir cómo ciertos mecanismos o elementos subyacentes fueron responsables de que cierto suceso físico o mental causara determinada sensación de dolor, y cómo éste a su vez fue causalmente responsable de otros sucesos y comportamientos. La pregunta es si existe la esperanza de lograr una explicación constitutiva completa de las sensaciones que nos permita acceder a estos mecanismos y que ilumine la naturaleza de la experiencia.

Nagel ("¿Cómo es ser un murciélago?"), McGinn ("¿Podemos resolver el problema mente-cuerpo?") y Jackson ("*Qualia* epifenoménicos" y "Lo que María no sabía") son todos (aunque en diferente medida) pesimistas al respecto. El problema reside esencialmente en que, al ofrecer explicaciones, lo hacemos desde un marco conceptual que es en cierto sentido *objetivo*, esto es, que no requiere una perspectiva o punto de vista particular sobre el objeto de estudio;[4] pero las sensaciones, y, en particular, su aspecto fenoménico es necesariamente *subjetivo*, esto es, requiere necesariamente un tipo de perspectiva o punto de vista. Consideremos primero la posición y argumentación de Nagel.[5]

Un relámpago puede conocerse de diversas maneras: unas son perceptuales y otras son meramente conceptuales. Puedo llegar a saber qué cosa es un relámpago al verlo caer, pero alguien que carezca de nuestro sistema perceptual puede también conocer nuestro concepto *relámpago* al saber que es una descarga eléctrica de cierto tipo.[6] Lo mismo sucede con el agua y el azúcar. Nuestros usos de los conceptos de *agua* y *azúcar* llevan

[4] El sentido de "objetivo" que nos interesa no es de lo que puede darse sin un punto de vista, sino sólo de aquello que puede darse independientemente de un tipo de punto de vista particular, esto es, de aquello que puede darse desde varios tipos de puntos de vista.

[5] El artículo de Nagel (publicado originalmente en 1974) inició la discusión contemporánea acerca de la conciencia fenoménica o *qualia*, de las cualidades fenoménicas de ciertos estados mentales. Pero el problema ciertamente surge mucho antes en la historia de la filosofía. El problema se puede trazar hasta Sexto Empírico (*Esbozos pirrónicos*), quien afirmó contra el atomismo de Epicuro que es imposible explicar cómo de un cúmulo de átomos puede surgir el placer. Leibniz (1765, véase el prefacio) también se preguntó cómo podían las sensaciones y el pensamiento surgir a partir de algo meramente mecánico. Como es bien sabido, Descartes (1637), también se enfrentó al problema e intentó resolverlo de manera no satisfactoria apelando a la glándula pineal como vínculo entre lo mental y lo físico.

[6] Somos neutrales sobre la cuestión de si los conceptos se individúan sólo por sus extensiones o por ellas y por la práctica asociada o sus usos. Para los fines de esta discusión, no mucho depende de esto. Si uno piensa, en una línea kripkeana, que los conceptos se individúan sólo en términos de sus extensiones, puede hacerlo sin afectar la problemática que aquí nos ocupa. Diría simplemente que alguien que carezca de mi sistema perceptual puede conocer mi concepto relámpago y no sólo su extensión. El que piensa que la práctica también es un elemento individuador, dirá que quien no comparte

asociados cierta fenomenología (por ejemplo, cómo se ven y cómo saben) y, por ello, cierto punto de vista sobre ellos. Pero, nuevamente, alguien que carezca de nuestro punto de vista sobre el agua o el azúcar por carecer de un sistema gustativo o visual como el nuestro, puede conocer el agua y el azúcar mediante una descripción de su composición química. Para ser comprendida en su totalidad, dicha descripción no requiere que uno tenga cierto aparato perceptual o cierta perspectiva o punto de vista sobre el agua o el azúcar. Sin embargo, la situación es muy distinta con las sensaciones.

Poseer el concepto de dolor o de sensación de rojo, a diferencia de la posesión de conceptos como *agua*, *azúcar* y *relámpago*, necesariamente requiere una perspectiva. La manera en que el agua o el azúcar o un relámpago se me presentan puede ser distinta de lo que la constituye. Así, por ejemplo, el agua se me presenta como transparente, típicamente inodora y con cierto sabor, pero está constituida por dos átomos de hidrógeno y uno de oxígeno. La posibilidad de esta diferencia entre cómo se me presenta algo y cómo es en realidad simplemente no surge en el caso de las sensaciones. La manera como se me presenta un dolor simplemente es ese dolor. No hay, en este caso, ninguna diferencia entre cómo se me presenta un dolor y cómo es ese dolor; en este caso no existe una distinción entre la apariencia y la realidad. Luego, si alguien ha de saber qué es un dolor, si siquiera ha de poseer el concepto *dolor*, deberá saber cómo se le presenta a un sujeto, cómo ese sujeto lo experimenta o siente. Y para saber esto, deberá, según Nagel, sentirlo. Asimismo, para dar cuenta de los dolores y demás sensaciones, el teórico deberá tomar en cuenta la manera como se le presentan al sujeto que las tiene, deberá hacerlo diciendo cómo las siente el sujeto, esto es, desde el punto de vista del sujeto. No hacerlo implicaría dar una explicación no satisfactoria de las sensaciones, una explicación que deja fuera lo esencial de las sensaciones.

El reto es, pues, dar cuenta de las sensaciones en términos que rescaten su aspecto fenoménico y, por ende, que tomen en cuenta la perspectiva que el sujeto tiene sobre ellas, esto

mi sistema perceptual sólo conoce la extensión de mi concepto, pero no mi concepto.

es, la perspectiva de la primera persona. El problema es que comúnmente las explicaciones se dan desde el punto de vista del observador o de la tercera persona, y no desde el punto de vista del sujeto. Ilustremos la cuestión asumiendo una perspectiva fisicalista de lo mental.

Un fisicalismo sostiene que lo mental es idéntico a lo físico, ya sea porque un tipo mental es idéntico a un tipo físico (fisicalismo tipo) o porque un suceso mental es idéntico a un suceso físico (fisicalismo caso). El fisicalismo tipo es poco defendido hoy en día, pero le sobrevive un fisicalismo caso, según el cual un mismo suceso o estado es mental y físico a la vez. El trabajo de Nagel está dirigido principalmente contra el fisicalismo tipo, pero, como veremos, sus consideraciones se aplican no sólo al fisicalismo tipo sino también al fisicalismo caso. El fisicalismo tipo sostiene que así como hay un tipo químico que corresponde con el tipo que hemos identificado perceptualmente como agua, dando como resultado enunciados de identidad como "agua = H_2O", de la misma manera hay un tipo físico que corresponde a un tipo mental como, por ejemplo, un dolor de cierta clase, que da como resultado enunciados de identidad como "el disparo de las fibras C = el dolor de tipo 1". Pero no importa qué caracterización física demos de un dolor, esta caracterización no servirá como análisis del concepto *dolor* de tipo 1, porque dejará sin explicación el aspecto fenoménico esencial a dicho concepto. Dejará sin explicación por qué ese tipo físico y no otro es idéntico al dolor de tipo 1 y cómo es que ese tipo físico puede ser responsable de cómo se siente el dolor. El problema es, pues, el siguiente:

> Si hay que defender el fisicalismo, debemos ofrecer una explicación física de las características fenomenológicas. Pero cuando examinamos su carácter subjetivo, tal resultado parece imposible de alcanzar. La razón es que todo fenómeno subjetivo está conectado esencialmente con un solo punto de vista, y parece inevitable que una teoría física objetiva abandone ese punto de vista. (Nagel, pp. 47–48 de este volumen.)

El famoso ejemplo del murciélago de Nagel sirve para ilustrar la importancia del aspecto subjetivo y el conflicto al que nos enfrentamos cuando intentamos explicar, desde un punto de

vista independiente, algo que depende de un punto de vista subjetivo.

Hemos dicho ya que lo esencial a la sensación de dolor o de rojo es la manera en que se siente o experimenta. Dado que la sensación es necesariamente sentida o experimentada, debe haber una criatura u organismo que sienta dicha sensación, esto es, la sensación debe pertenecer a algo que sea consciente. Y si hay algún organismo que tenga experiencias conscientes entonces, según Nagel, hay algo que es cómo es ser ese organismo para ese organismo, y viceversa. El cómo es ser una criatura para ella nos da el punto de vista subjetivo necesario para cualquier experiencia, y es lo que Nagel llama "el carácter subjetivo de la experiencia".

Ciertamente, un murciélago tiene experiencias. Según Nagel, esas experiencias son conscientes en el sentido de que hay algo que es cómo se siente tal experiencia para el murciélago y, por ello, hay algo que es cómo es ser un murciélago para el murciélago.[7] Pero, ¿podríamos nosotros saber cómo es ser un murciélago? Parte de lo que constituye la manera de ser de un organismo son las experiencias que tiene,[8] y éstas están determinadas, al menos en parte, por los aparatos perceptuales de ese organismo.[9] La mayoría de las experiencias perceptuales de los murciélagos provienen principalmente de su aparato sonar, por medio de ecolocalización. Los murciélagos emiten ciertos sonidos y la detección de sus reflejos les da la información necesaria para maniobrar en el mundo. Luego, si queremos

[7] El ejemplo supone que los murciélagos tienen cierta conciencia de sí mismos. Esto no es fácil de aceptar. Sin embargo, el ejemplo podría modificarse sustituyendo a los murciélagos por seres autoconscientes (e imaginarios) con un sonar semejante al de los murciélagos. En lo que sigue respetaremos la formulación original del ejemplo de Nagel.

[8] Es difícil dar una lista exhaustiva de lo que constituye o determina cómo es ser un organismo para ese organismo. Las experiencias perceptuales son parte de ello, pero también lo son las capacidades que el organismo tiene para interactuar con el mundo y, en nuestro caso, ciertas creencias, deseos e incluso ciertas tendencias psicológicas como las tendencias a la depresión o la hiperactividad.

[9] Recordemos que hay experiencias que no son perceptuales, como el dolor y la sensación de hambre, entre otras. Ambas atribuibles, según Nagel, a los murciélagos.

saber cómo es ser un murciélago, es necesario que sepamos cómo es tener una experiencia ecolocalizadora. El problema es que carecemos de un sentido perceptual semejante al sonar, y por ello de experiencias ecolocalizadoras necesarias para saber cómo es ser un murciélago.

¿Podríamos imaginar cómo es tener experiencias ecolocalizadoras para llegar a saber cómo es ser un murciélago? Podría imaginarme que tengo los rasgos de un murciélago (que tengo alas, que vuelo al atardecer, que como fruta, que tengo muy mala visión), pero esto no constituiría imaginar cómo es ser un murciélago *para un murciélago*, sino cómo sería *para nosotros* ser un murciélago. El problema es que nuestras capacidades imaginativas están restringidas por nuestras experiencias y recursos mentales, los cuales son insuficientes para imaginarnos la fenomenología de un murciélago, esto es, cómo es ser un murciélago *para un murciélago*.[10] En particular, sin acceso a la experiencia fenoménica ecolocalizadora de un murciélago, no tenemos acceso al carácter subjetivo de esa experiencia. Podemos ofrecer todo tipo de descripciones de esas experiencias (por ejemplo, podemos describirlas como experiencias tridimensionales auditivas), e incluso explicaciones causales de los mecanismos que hacen funcionar las experiencias ecolocalizadoras, pero esto no bastará para saber cómo es tener esas experiencias, y por ello, tampoco para saber cómo es ser un murciélago. La fenomenología de una experiencia ecolocalizadora, y por ello, la experiencia ecolocalizadora misma y el cómo es ser un murciélago para un murciélago es inefable para nosotros. Según Nagel, hay entonces hechos que son esencialmente subjetivos, esto es, que dependen del punto de vista del organismo que los tiene. En algunos casos, como los de las experiencias del murciélago, constituyen hechos cuyo conocimiento y comprensión escapan a las capacidades humanas, en particular, a los conceptos humanos.

En nuestro propio caso, nuestras experiencias no son inefables para nosotros, pero sí son esencialmente subjetivas; a saber, dependen de cierta clase de punto de vista o perspectiva.

[10] Véase la n. 7.

Como en el caso de las experiencias del murciélago en relación con nosotros, para una criatura que careciera de nuestro aparato perceptual nuestras experiencias serían inefables, pues carecería del tipo de perspectiva necesaria para saber cómo se sienten esas experiencias. La dificultad para el fisicalismo reside en que si los hechos de la experiencia son hechos acerca de cómo es para el organismo que los experimenta experimentarlos, esto es, si sólo son accesibles desde un punto de vista, entonces resulta un misterio cómo una explicación física objetiva o independiente de ese punto de vista puede dar cuenta de las experiencias. Recordemos que las explicaciones físicas (composicionales y causales) como las de la solubilidad del azúcar en el agua o de la naturaleza de un relámpago son explicaciones que son, en principio, accesibles por criaturas que no comparten nuestros aparatos perceptuales. Esto, como el caso del murciélago nos ha mostrado, no sucede en el caso de las propias experiencias. Decir que el disparo de las fibras C constituye o es simplemente un dolor de tipo 1 no nos dice nada acerca de la fenomenología de ese dolor, de por qué el disparo de esas fibras constituye la manera en que el sujeto siente el dolor. Poniéndolo de otro manera, el aspecto relacional dependiente de un punto de vista que es esencial a la experiencia es ignorado en la explicación fisicalista. Luego, queda un hueco explicativo por cerrar.[11]

Este hueco existe incluso con el fisicalismo caso. El fisicalismo caso, a diferencia del fisicalismo tipo, sostiene que no hay identidad entre tipos mentales y tipos físicos, pero que todo suceso mental es un suceso físico. Así, el suceso que constituye mi dolor en la rodilla es un suceso que tiene propiedades físicas, específicamente, neurofisiológicas. Un fisicalismo caso, como tal, no está comprometido a elucidar la relación entre las propiedades mentales de un suceso y las físicas más allá de decir que cada vez que hay un suceso con propiedades mentales, éste tendrá propiedades físicas. Sin embargo, esto no ayuda a cerrar el hueco explicativo. Ha habido intentos de decir algo más sobre la dependencia que lo mental tiene de lo físico. En particular, se ha dicho que existe una relación de *superveniencia* que lo

[11] Sobre el hueco explicativo, véase Levine 1983.

mental tiene de lo físico.[12] Hay diferentes maneras de entender esta relación de superveniencia,[13] pero la idea general se puede expresar así: si un objeto, suceso o mundo E′ difiere en sus propiedades mentales de un objeto, suceso o mundo E″, entonces E′ y E″ difieren en sus propiedades físicas. Sin embargo, no hay nada en la relación de superveniencia que ayude a cerrar el hueco explicativo.[14] El hueco subsiste porque la relación de superveniencia que establece ciertas correlaciones necesarias no puede ella sola decirnos cómo (por ponerlo en términos de Nagel) de un conjunto de hechos objetivos surge un hecho subjetivo, no puede ofrecernos una explicación constitutiva de un hecho subjetivo a partir de hechos objetivos.

Pero aceptar la existencia de este hueco explicativo no es aún concluir que todo fisicalismo es falso,[15] o que no podremos algún día cerrar ese hueco. La conclusión de Nagel no consiste en negar el fisicalismo, no es siquiera una conclusión ontológica. Adopta más bien una posición epistemológica, en particular, una posición agnóstica: dado el estado de nuestro conocimiento actual en ciencia y filosofía no sabemos cómo el fisicalismo podría ser verdadero, cómo podría cerrarse el hueco explicativo y cómo podríamos lograr una explicación fisicalista (causal y constitutiva) satisfactoria del carácter subjetivo de la experiencia. Quizá en el futuro podamos desarrollar un vocabulario[16] que nos permita entender cómo de la actividad cerebral puede surgir la conciencia fenoménica.

A diferencia de Nagel, McGinn ("¿Podemos resolver el problema mente-cuerpo?") arguye que nunca podremos cerrar ese hueco explicativo porque comprender la manera en que éste pudiera cerrarse trasciende nuestras capacidades cognoscitivas;

[12] Ejemplos de esto son Davidson 1970, Kim 1993, Fodor 1987, entre otros.

[13] Véase Kim 1984.

[14] Nagel tampoco cree que un análisis funcionalista o causal de lo mental logre cerrar el hueco explicativo. Sobre el funcionalismo volveremos más adelante.

[15] A partir del carácter esencialmente subjetivo de lo mental y dentro del marco de su teoría de designadores rígidos, Kripke (1972, véase su conferencia III) extrae la conclusión de que el fisicalismo es falso.

[16] Nagel habla incluso de la posibilidad de desarrollar una fenomenología objetiva, lo que a primera vista parece una contradicción en términos.

en palabras de McGinn, nuestras mentes están *cognoscitivamente cerradas* a la(s) teoría(s) necesaria(s) para cerrar dicho hueco. Un tipo de mente M está *cognoscitivamente cerrada* a una propiedad P (o teoría T) si y sólo si los procedimientos para formar conceptos que están a la disposición de M no pueden extenderse a una aprehensión de P (o una comprensión de T). McGinn argumenta que existe una propiedad P que el cerebro instancia, en virtud de la cual el cerebro es la base para la conciencia, y, correspondiente a esto, que existe una teoría T que refiere a P y que explica completamente la dependencia de los estados conscientes en los estados cerebrales. Así, de conocer T, tendríamos una solución al problema mente-cuerpo. Desgraciadamente, según McGinn, T está cognoscitivamente cerrada para nosotros. Si bien el fracaso a lo largo de la historia de la filosofía y (más recientemente) de la ciencia en explicar cómo surge la conciencia a partir de la actividad cerebral muestra que hay una dificultad seria para resolver el hueco explicativo, esto no permite todavía concluir que simplemente no hay esperanza de cerrar dicho hueco. Sin embargo, bastaría mostrar que no podríamos formar los conceptos pertinentes de P para afirmar que nuestras mentes están cognoscitivamente cerradas a P.

Ciertamente, la introspección es el medio aparentemente más adecuado para acceder a la fenomenología de la conciencia,[17] pues tenemos acceso inmediato a ella. Por ejemplo, para saber cómo se siente mi dolor no tengo más que sentirlo o introspectar dicha sensación y sabré cómo es mi dolor. Pero la introspección no es el medio ideal para acceder al vínculo entre lo mental y lo físico, pues no tenemos acceso a lo físico por medio de la introspección. La percepción directa del cerebro tampoco sirve como medio para llegar a conocer P o para formar un concepto de P. La razón es que la percepción es esencialmente un medio para conocer objetos espaciales, pero la conciencia no es algo que se ubique en el espacio o que tenga partes espaciales. Podemos pensar, no obstante, que el concepto pertinente de P será un concepto teórico que se introduce para explicar ciertos fenómenos observables del cerebro, esto es, que se introduce

[17] Como veremos en la sección 4, Dennett ("Quinear los *qualia*") cuestiona que esto sea cierto.

mediante un argumento a la mejor explicación. Éste sería un camino más prometedor si no fuera porque la introducción de conceptos teóricos en las ciencias físicas está constreñida por un principio de *homogeneidad*.[18] Dicho principio establece que, en la introducción de conceptos teóricos a partir de la observación de fenómenos físicos, estos conceptos teóricos serán conceptos físicos y no mentales. Según McGinn, esto significa que no podemos acceder a la conciencia a partir de las ciencias físicas y, por tanto, que la conciencia está cognoscitivamente cerrada frente a la introducción de conceptos vía inferencias a la mejor explicación a partir de datos perceptuales acerca del cerebro.

Ciertamente, gran parte de los conceptos que nos formamos están ligados con la percepción y la introspección, pero no todos. Nuestros conceptos matemáticos, al parecer, no lo están. McGinn cree que una mente que obtuviera la mayoría de sus conceptos como nosotros obtenemos los conceptos matemáticos podría (quizás) obtener un concepto de P y comprender la teoría T correspondiente. Dicha mente tendría que poder pensar sobre el cerebro y la conciencia fenoménica de modos que prescindieran de los medios perceptuales y de la introspección. McGinn sugiere que dicha mente podría pensar sobre el problema de manera *a priori*, semejante a la manera en que pensamos en números. Pero dado que nosotros no podemos formarnos conceptos del cerebro y de la conciencia independientemente de la percepción y la introspección, según McGinn, nuestras mentes están cognoscitivamente cerradas a P.

2. *El argumento del conocimiento*

En la misma línea de Nagel, aunque con pretensiones de llegar a una conclusión ontológica sobre el fisicalismo, Jackson ("*Qualia* epifenoménicos" y "Lo que María no sabía") presenta lo que se ha denominado el "argumento del conocimiento". Este argu-

[18] Este principio de homogeneidad se encuentra presente de alguna manera en las discusiones del fisicalismo caso. Davidson (1970), Crane (1991), y Papineau (1990 y 1991) (por mencionar a algunos) hablan de que lo físico constituye un sistema cerrado, esto es, un sistema que no admite elementos no físicos, entre éstos, los mentales.

mento, tan discutido en la literatura filosófica reciente, intenta mostrar que el fisicalismo es falso.

En el primer artículo, empieza afirmando la validez del argumento más común contra el fisicalismo: que nada de lo que se pueda decir físicamente captura, por ejemplo, el olor de una rosa. Pero, si el fisicalismo es correcto, entonces todo es capturable en términos físicos. Por lo tanto, el fisicalismo es incorrecto. Sin embargo, considera que este argumento es muy débil en el contexto de la polémica alrededor del fisicalismo, ya que sus premisas no son intuitivamente obvias para muchos filósofos. Por ello, se propone presentar un argumento contra el fisicalismo cuyas premisas sean obvias para la mayoría: el argumento del conocimiento.

El argumento se da en dos variantes. Ambas parten de experimentos de pensamiento en los que ciertos individuos, que nunca han experimentado una determinada experiencia perceptual o una determinada sensación, poseen, sin embargo, toda la información física pertinente (conocimiento científico completo) sobre esa experiencia perceptual o esa sensación. Aquí sólo nos ocuparemos de la segunda variante, que ha sido la más discutida.

Se trata de una brillante científica, María, que sabe todo lo que hay que saber en términos físicos (tiene conocimientos científicos completos) de la naturaleza física del mundo, incluyendo todo lo que hay que saber acerca de los hechos causales relacionales y los papeles funcionales. Conoce, por supuesto, todo acerca del proceso visual y las experiencias de los colores. Sin embargo, como está confinada a un mundo blanco y negro (que incluye que vea su cuerpo en blanco y negro) y estudia con libros y monitores de televisión en blanco y negro, nunca ha experimentado sensaciones de colores como rojo o verde. Un día sale de su habitación y ve por primera vez un tomate. Según Jackson, María aprende algo nuevo, a saber, cómo se ven los tomates maduros. Pero, ¿cómo es esto posible si, por hipótesis, ella ya sabía todo lo físicamente pertinente acerca de las experiencias de rojo? Entonces, si el fisicalismo es la tesis de que el mundo es enteramente físico y el conocimiento físico es el conocimiento completo, esta tesis es falsa, puesto que María, al

salir de su cuarto aprende algo que no sabía, es decir, cómo se siente ver algo rojo.

En "Lo que María no sabía", Jackson aclara algunos aspectos de su argumento y lo defiende de varias objeciones. Explica cómo en su argumento no está en cuestión la intensionalidad del conocimiento, ni que éste dependa de los poderes imaginativos de María. También compara su argumento con el de Nagel y afirma que el objetivo de este último es mostrar que ninguna cantidad de información física (que es accesible desde muchos puntos de vista) nos puede hacer entender cómo es ser un murciélago. Esto sólo puede entenderse desde el punto de vista del murciélago, el cual es muy distinto al nuestro, por lo que los seres humanos no podemos saber y ni siquiera imaginar cómo es ser un murciélago. Según Jackson, a Nagel le preocupa el problema de extrapolar el conocimiento de una experiencia con el de otra, y el problema de imaginar una experiencia que no es familiar, a partir de una que sí lo es. Así, nos dice Jackson, el argumento de Nagel no es realmente un argumento contra el fisicalismo, dado que el fisicalismo no afirma nada en especial sobre los poderes imaginativos o de extrapolación de los seres humanos. En cambio, con su argumento del conocimiento, Jackson subraya el hecho de que hay algo de la experiencia misma que el fisicalismo deja de lado, porque al tener por primera vez una experiencia particular, sabemos algo que antes no sabíamos de la persona que tenía la experiencia, aunque sigamos ignorando cómo es ser esa persona (por ejemplo, cómo es ser María). Así, de acuerdo con Jackson lo que le faltaba a María es *conocimiento acerca de las experiencias de los otros*. Cuando María sale y experimenta por primera vez cómo es ver un tomate, se da cuenta de que la concepción que tenía de la vida mental de los *otros* (y de ella) estaba muy empobrecida ya que no sabía cómo era ver los colores. Dado que ella sabía todo lo físico acerca de la vida mental de los otros, lo que aprende tras su liberación es un hecho no físico acerca de ella y éste es el problema que no puede manejar el fisicalismo.

El argumento ha sido discutido por numerosos autores, entre ellos Churchland[19] y Lewis. Aquí sólo hemos incluido el

[19] De hecho, "Lo que María no sabía" se escribió en respuesta al artículo

artículo de Lewis ("Lo que enseña la experiencia"), en el que argumenta contra Jackson sosteniendo que lo que María adquiere cuando experimenta por primera vez una sensación de rojo no es conocimiento proposicional sino una capacidad recognoscitiva. Cuando Jackson suponía que María conocía todo lo físico acerca de la visión de los colores, lo que ella tenía era conocimiento proposicional. En cambio, aprender cómo es cierta experiencia es adquirir nuevas habilidades para recordar, imaginar y reconocer. Lo que María adquiere cuando es liberada es cierta habilidad representacional o imaginativa; es un saber *cómo* en vez de un conocer *que*. Si esto significa adquirir nueva información, entonces hay buenas razones para pensar que esta información nueva no es un tipo especial de información no física como pretende Jackson. Por ello, concluye Lewis, el hecho de que uno no pueda conocer cómo es una nueva experiencia a menos que uno la experimente, no debe tomarse como una medida de los límites de la ciencia o de la metafísica de la mente.

Lewis parece tener razón al decir que María adquiere ciertas habilidades o capacidades nuevas, por ejemplo, reconocer, imaginar o recordar lo que se siente ver rojo. Sin embargo, la pregunta es si María adquiere algo más que habilidades. ¿Conoce algo nuevo acerca de la percepción de los colores sólo después de salir del cuarto? ¿Conoce un hecho acerca de cómo aparece el rojo, cómo se siente ver rojo, o quizá, también, *que* los otros ven el rojo de un cierto modo? ¿Es esto conocer de un modo nuevo algo que proposicionalmente ya conocía?

Esta última es la posición de Churchland (1985, 1989b), quien niega que María conozca nuevos hechos cuando sale del cuarto, pero admite que los conoce de un nuevo modo. Tye (1986) también afirma que María obtiene un nuevo conocimiento sólo en el sentido de que ahora tiene, mediante la introspección, un nuevo modo de acceso a aquello que ya conocía antes por inferencia. Está ahora en un nuevo estado epistémico debido a que ha adquirido un nuevo sistema de representación. Como pode-

de Churchland (1985). Véanse este artículo y Churchland 1989b para una discusión del argumento del conocimiento.

mos ver, no se ha dicho todavía la última palabra en relación con este argumento.[20]

3. *Funcionalismo,* qualia *ausentes e inversiones*

Hasta ahora nos hemos enfocado en problemas que la conciencia fenoménica trae al fisicalismo, pero Nagel mismo pretendía que sus consideraciones se aplicaran también a teorías funcionalistas. En efecto, se aplican las mismas consideraciones; pero además el funcionalismo da lugar a otros problemas relacionados con la conciencia fenoménica, como son los de las hipótesis de los *qualia* ausentes y del espectro invertido.

El artículo de Block y Fodor ("Lo que no son los estados psicológicos") no discute como tema principal la conciencia fenoménica, aunque critica ciertas teorías de identidad tipo entre lo mental y lo físico, así como una clase de funcionalismo. Sin embargo, lo más importante para esta antología es que presenta un reto, olvidado hasta la aparición de este artículo, al que el funcionalismo se enfrenta con los *qualia*.[21] El funcionalismo clásico es una teoría de identidad tipo entre los estados (o sucesos) mentales y sus papeles causales o funcionales. Así, si e_1 y e_2 tienen las mismas causas y efectos, entonces, según el funcionalismo e_1 es idéntico a e_2. Block y Fodor caracterizan lo que denominan la teoría de identidad de los estados funcionales (en adelante TIEF). Ésta supone que un organismo al cual se aplican predicados psicológicos puede describirse como un autómata probabilístico, cuya descripción se entiende como un enunciado verdadero de que dicho organismo posee diferentes estados relacionados entre sí y con las entradas y salidas del organismo mediante las probabilidades de transición especificadas en la matriz o tabla de máquina. Según la TIEF, existe una descripción única óptima de cualquier organismo que satisface predicados psicológicos conforme a la cual cada estado psicológico suyo

[20] Véase Loar 1990 y van Gulick 1993.

[21] Decidimos que más que presentar sólo una parte del artículo, la parte que tiene que ver con las sensaciones y la conciencia fenoménica, era preferible incluirlo completo para dejar a criterio del lector la selección de lo pertinente en la discusión del funcionalismo y la conciencia fenoménica, y también porque el artículo no ha sido publicado anteriormente en español.

es idéntico en tipo a uno de sus estados de tabla de máquina relativos a esa descripción. Así, dos organismos están en estados psicológicos del mismo tipo si y sólo si están en el mismo tipo de estado de tabla de máquina y dos estados de tabla de máquina son idénticos si y sólo si se relacionan de la misma manera con otros estados de tabla de máquina y con los estados de los mecanismos de entrada y salida. Uno de los problemas a los que se enfrenta la TIEF es el de dar cuenta de estados mentales que tienen un carácter cualitativo esencial. Dos argumentos claros contra la TIEF en particular, y el funcionalismo en general, tienen que ver con las ya mencionadas hipótesis del espectro invertido y de los *qualia* ausentes o de lo que más recientemente se ha caracterizado como la posibilidad de zombis.

Supongamos que tres sujetos S', S'' y S''' son idénticos en su arreglo funcional interno. Supongamos que S' tiene experiencias con una fenomenología semejante a la nuestra. Ahora bien, el argumento de los *qualia* ausentes se basa en la posibilidad de que haya dos sujetos funcionalmente idénticos (o cuyas descripciones óptimas de sus tablas de máquina sean idénticas), S' y S'', que difirieren en que uno tiene fenomenología y el otro no. En otras palabras, en que S' experimente *qualia* mientras que S'', quien tiene las mismas creencias, deseos y demás estados mentales con las mismas relaciones causales que S', no experimente ningún *quale*.[22] El caso de la hipótesis del espectro invertido supone que puede haber dos sujetos funcionalmente idénticos, o cuyas descripciones óptimas de sus tablas de máquina sean las mismas, S' y S''', pero cuyos espectros visuales estén invertidos uno con respecto del otro. Así, S' tendrá la fenomenología que nosotros tenemos cuando observamos la hierba como verde, pero S''' tendrá la fenomenología invertida con respecto a S' y, por ende, a nosotros. Cuando S''' ve la hierba tendrá la fenomenología que nosotros tenemos cuando vemos un objeto rojo, y así sucesivamente para las diferentes experiencias de los colores.

¿Cómo puede responder el funcionalista a argumentos que pretendan mostrar que el funcionalismo es falso utilizando la

[22] Para otras versiones de la posibilidad de *qualia* ausentes y un arreglo funcional idéntico, véase Block 1980c.

posibilidad de los *qualia* ausentes y del espectro invertido? Primero, si el funcionalismo se propone, como tradicionalmente se ha hecho, como una teoría de identidad tipo, entonces debería o bien explicar cómo estos dos supuestos casos en realidad no son posibles o bien que el que lo sean no representa problema alguno para el funcionalismo. Sin embargo, si el funcionalista sostiene una tesis menos estricta de la relación que se da entre lo mental y los papeles funcionales, entonces podrá aceptar la posibilidad de zombis y de espectros invertidos, aunque deberá ofrecer una explicación constitutiva y no funcional de por qué y cómo ciertos estados mentales tienen fenomenología.[23]

Así, en tanto teoría de identidad tipo, la TIEF está obligada a dar respuesta a la posibilidad del espectro invertido y de los *qualia* ausentes. Block y Fodor, en un intento de defender la TIEF, sostienen que la posibilidad de un espectro invertido, aunque es una posibilidad lógica, no es una posibilidad nomológica según las leyes de nuestro mundo actual. Ciertamente, hoy en día y dados los avances de nuestras ciencias, no podemos apreciar cómo es que el espectro invertido no es empíricamente posible. Sin embargo, el defensor del funcionalismo diría que cuando tengamos una teoría acabada de los papeles funcionales de los estados mentales veremos cómo la hipótesis del espectro invertido no es posible. Lo mismo diría frente a la posibilidad de los *qualia* ausentes: aunque sea una posibilidad lógica, no constituye una posibilidad empírica dadas las leyes de nuestro mundo.

Aunque estas respuestas son atractivas, no son totalmente aceptables para una teoría de identidad como lo es la TIEF. No lo son al menos si predicados de experiencias como "tener un dolor del tipo 1" y predicados funcionales como "tener cierto papel funcional *F*" son designadores rígidos, esto es, frases que designan (si acaso) la misma clase (o el mismo objeto cuando se

[23] La relación en cuestión no podrá ser de superveniencia fuerte, sino sólo de superveniencia débil, *i.e.*, una relación que se sostiene sólo para nuestro mundo, y no con respecto a cualquier otro mundo posible como sería necesario con una relación de superveniencia fuerte.

trata de términos singulares) a través de todo mundo posible.[24] Kripke (1972) ha argumentado convincentemente que los enunciados de identidad en los que los términos al flanco del signo de identidad son designadores rígidos expresan una identidad necesaria, *i.e.*, que se sostiene a través de todos los mundos posibles. Luego, si "tener un dolor del tipo 1" y "tener cierto papel funcional *F*" son designadores rígidos,[25] la TIEF debe ofrecer otra respuesta frente a las posibilidades de los *qualia* ausentes y del espectro invertido. Nuevamente, tendrá que negar que existan esas posibilidades o decir por qué no le deben preocupar al funcionalista.

Shoemaker ("Funcionalismo y *qualia*") intenta defender el funcionalismo frente a argumentos basados en estas posibilidades lógicas. Su estrategia es, por un lado, negar la posibilidad de los *qualia* ausentes sobre la base de la imposibilidad de saber si se dan o no y, por otro lado, aceptar la posibilidad de los *qualia* invertidos intentando mostrar que se puede admitir dentro de un tipo de funcionalismo.

Suponiendo que *S''* sea funcionalmente idéntico a *S'*, no sólo exhibirá la misma conducta, sino que ambos tendrán los mismos estados mentales, excepto que *S''* no tendrá ninguna fenomenología. Así, *S''* tendrá cuasidolores o dolores *ersatz*, pues se encontrará en estados funcionalmente idénticos a los estados de dolor en los que está *S'*, pero sus estados carecerán de un cómo se siente. Uno de los aspectos respecto a los cuales serán idénticos será en cuanto a las *creencias cualitativas* que dichos estados de dolor (o cuasidolor) causan. Éstas son creencias acerca de sus estados de dolor (o cuasidolor) en el sentido de que se está teniendo un dolor. Si esto es el caso, entonces ambos, *S'* y *S''*, creerán que tienen un dolor y lo reportaran así. La evidencia a la que recurrimos para saber qué estados tiene un sujeto distinto de nosotros es observacional: observamos su comportamiento lingüístico y no lingüístico y, a través de él, atribuimos estados mentales. En nuestro propio caso, sin embargo, recurrimos a la introspección. Ciertamente, no hay nada

[24] Por cuestiones de simplicidad, estamos ignorando aquí los mundos posibles en los que no existen las clases.

[25] Este antecedente no es obvio, y es algo que tendría que discutirse con mayor detalle. Ésta es una discusión que dejamos de lado aquí.

en el comportamiento de *S''* que nos permita saber, desde la perspectiva de la tercera persona, que *S''* carece de fenomenología. Pero tampoco podría *S''* mismo desde la primera persona saber que carece de fenomenología por medio de la introspección. Él tendrá la creencia cualitativa de que tiene un dolor con cierta fenomenología en tanto sea funcionalmente idéntico a *S'*. Y si *S''* no puede saber por medio de la introspección que no es un zombi, entonces tampoco puede saberlo *S'*, y por ello, tampoco nosotros mismos podemos saber que no somos zombis. Según Shoemaker, aceptar la posibilidad de los *qualia* invertidos sería comprometerse con un escepticismo intolerable acerca del conocimiento de nuestros estados mentales desde la perspectiva de la primera persona. La estrategia argumentativa de Shoemaker aquí se basa en una reducción al absurdo: la posibilidad lógica de los *qualia* ausentes nos lleva a consecuencias epistémicas intolerables, por lo cual debemos rechazar la posibilidad lógica de los *qualia* ausentes.[26]

¿Cómo podríamos saber que se dan casos de espectros invertidos? Hay dos maneras en que podemos hablar de *qualia* invertidos. Uno es el caso *inter*subjetivo como el de *S'''* respecto de *S'* y otro es el caso *intra*subjetivo, según el cual un mismo sujeto *S''''* sufre una inversión de su espectro. En el primer caso, como en el caso de los *qualia* ausentes, simplemente no tendríamos ningún tipo de evidencia (ni conductual ni introspectiva) que nos permitiera saber (o le permitiera saber a *S'''*) que el espectro de *S'''* está invertido con respecto al nuestro, pues su comportamiento será *ex* hipótesis igual al de *S'*, cuyo espectro no está invertido. Sin embargo, en el caso intrasubjetivo en el que el sujeto mismo sufre una inversión de su espectro visual podrá él mismo, mediante la introspección, detectar el cambio sufrido. Luego, a diferencia del caso de los *qualia* ausentes, existe la posibilidad de llegar a saber que se da un caso de inversión de *qualia*. Cuando *S''''* sufra la inversión de su espectro, podrá por medio de la introspección pensar (e incluso decir) que el cielo se ve hoy como ayer se veían los plátanos y la

[26] Para una versión desarrollada de este argumento, véase el artículo de Shoemaker incluido en esta antología y Shoemaker 1981. Y para algunas discusiones del argumento, véanse Block 1980a y Conee 1985.

hierba se ve hoy como ayer se veían las fresas, etc. Dados estos juicios no queda claro que en este caso *S''''* tenga el mismo arreglo funcional que el día anterior, pues estas nuevas creencias serán efectos de los estados perceptuales con nuevos *qualia* asociados. Sin embargo, al cabo de un tiempo, el sujeto podría en principio tener un arreglo funcional idéntico al que tenía antes de la inversión del espectro, con los *qualia* de sus estados mentales (visuales) invertidos con respecto a como los tenía antes.

Si se acepta la posibilidad lógica de esta inversión intrasubjetiva, entonces una teoría de identidad tipo funcionalista como la TIEF no se sostendrá para el carácter cualitativo de los estados mentales. Sin embargo, esto no conlleva que el funcionalismo no pueda distinguir estados mentales cualitativos de estados mentales no cualitativos. Shoemaker sugiere que lo puede hacer a través de las relaciones de similitud y diferencia cualitativa. *S'* puede comparar (como cualquier otro sujeto) el carácter cualitativo de su estado al ver la hierba con aquel de su estado al ver el cielo, y esto da lugar a creencias acerca de las diferencias en el mundo de los objetos percibidos: las experiencias visuales del cielo y de la hierba pueden llevar a que *S'* se forme la creencia de que el cielo y la hierba no son del mismo color. Asimismo, dos estados visuales simultáneos que tienen un carácter cualitativo semejante normalmente llevarán a un sujeto a formarse la creencia de que los objetos en el mundo que ve tienen colores semejantes y, más aún, a formarse la creencia de que las experiencias en cuestión tienen un carácter cualitativo semejante. Utilizando esta caracterización de semejanza cualitativa de dos estados, Shoemaker da una definición de qué es que un estado tenga un carácter cualitativo: el estado de un sujeto tiene un carácter cualitativo si y sólo si es cualitativamente semejante a otro estado de ese sujeto (o distinto de él). Con la caracterización funcional de la semejanza cualitativa, Shoemaker cree que podemos caracterizar funcionalmente clases de estados mentales con carácter cualitativo, esto es, con fenomenología. Aun cuando esta caracterización no nos dé las diferencias entre *S''''* antes de su inversión y *S''''* después de su inversión, sí nos dará las diferencias entre estados mentales cualitativos y estados mentales no cualitativos como lo son las creencias. La diferencia radicaría en que en la caracterización

de los estados mentales cualitativos será necesario apelar a la semejanza cualitativa, la cual es definible de manera funcional.[27]

Ciertamente, la defensa de Shoemaker del funcionalismo no muestra, ni pretende mostrar, que una teoría de identidad tipo funcionalista como la TIEF es correcta. Tampoco, si la inversión de *qualia* es posible, permite que el arreglo funcional de los estados mentales de un sujeto puedan cerrar el hueco explicativo, esto es, puedan mostrar cómo de ciertos arreglos funcionales surge el carácter cualitativo o los *qualia*. No obstante, esto no significa que desechemos la importancia de una teoría funcionalista de la mente para los otros estados mentales y para entender las relaciones causales y epistémicas que tienen estados cualitativos con otros estados mentales como las que tienen con las creencias cualitativas. Sólo significa que si la inversión de *qualia* es posible, no será el arreglo funcional el que explique cómo se siente cierta experiencia.

4. *Características de los* qualia *y la posibilidad de un eliminativismo*

Hasta ahora hemos visto que la conciencia fenoménica es esencialmente subjetiva en tanto que comprenderla y conocerla depende esencialmente de la perspectiva del sujeto, y en este sentido es relacional. También hemos visto que si la posibilidad del espectro invertido existe, entonces el carácter cualitativo o la conciencia fenoménica asociada a un estado no es capturable en términos de las relaciones causales o funcionales de ese estado. En este otro sentido podríamos argumentar que el carácter cualitativo de una experiencia no es relacional sino más bien intrínseco. Así, en cierto sentido, dicho carácter es relacional y, en otro, es intrínseco. No hay contradicción aquí. Pero, ¿no habrá acaso en la noción de carácter cualitativo o *qualia* una suerte de incoherencia? Ésta es a grandes rasgos la sugerencia de Dennett ("Quinear los *qualia*"), quien con base en una suer-

[27] Para una versión más desarrollada de este argumento, véase, además del artículo incluido en esta antología, Shoemaker 1982 y para discusiones y réplicas a este argumento y al anterior, véase Levine 1988. Una defensa interesante del funcionalismo frente a la posibilidad de *qualia* invertidos y de *qualia* ausentes se encuentra en Chalmers 1995.

te de "incoherencia" atribuida al concepto de *quale* argumenta que los *qualia* simplemente no existen.[28]

Dennett afirma que en la literatura filosófica tradicional sobre conciencia fenoménica o *qualia* se ha sostenido que los *qualia* tienen esencialmente cuatro propiedades: son inefables,[29] son intrínsecos, son privados y son aprehensibles de manera directa o inmediata en la conciencia. Sostiene que aunque hay intentos de argumentar en contra de una u otra propiedad como propiedades de los *qualia*, inevitablemente la propiedad que se pretende desechar resurge de una u otra manera. Dennett considera esto como indicio de que realmente las cuatro propiedades atribuidas a los *qualia* son esenciales al concepto. Sin embargo, Dennett cree que estas cuatro propiedades llevan a una suerte de "paradoja" —ciertamente, no a una paradoja formal, pues ésta supondría que tenemos un concepto claro de *qualia*, que él cree que no tenemos. A partir de estas "paradojas", Dennett atribuye "incoherencia" al concepto preteórico de *qualia*.

Aquí sólo consideraremos esquemáticamente una de las cuestiones que surgen con el supuesto de que los *qualia* son directamente accesibles en la conciencia, y sobre los cuales tenemos autoridad de primera persona, así como dos críticas a la estrategia general de Dennett.

[28] En contraste con las conclusiones epistemológicas de Nagel y McGinn sobre los *qualia* y el problema mente-cuerpo, Dennett, como Jackson, asume una posición ontológica. Sin embargo, la suya es distinta de la de Jackson, a saber, que los *qualia* simplemente no existen. Presumiblemente, sin ellos el problema mente-cuerpo se vuelve más tratable.

[29] En la sección 1 dijimos que Nagel consideraba que la conciencia fenoménica era inefable en cierto sentido, a saber, que para nosotros, por ejemplo, las experiencias fenoménicas ecolocalizadoras de los murciélagos nos eran inefables porque no podíamos compartir su punto de vista. Sin embargo, las experiencias humanas no nos eran inefables pues, dada nuestra neurofisiología y mecanismos perceptuales, en principio tenemos experiencias del mismo tipo y compartimos los conceptos de nuestros *qualia*. El sentido de inefabilidad del que Dennett habla en "Quinear los *qualia*" es más extremo. Supone que mis experiencias son inefables para cualquier otro ser humano, pues no puedo comunicarle exactamente cómo las siento. En "*Qualia* epifenoménicos", Jackson mismo considera que Nagel requeriría esta inefabilidad más extrema (véase la discusión en la sección 2), pero nosotros dejamos aquí de lado la pregunta de si la inefabilidad considerada por Nagel en "¿Cómo es ser un murciélago?" nos lleva a esta inefabilidad más extrema.

Por un lado, tenemos la intuición *prima facie* de que los *qualia* de nuestras experiencias nos son directamente accesibles en la conciencia y respecto de los cuales tenemos autoridad de primera persona y cierta infalibilidad. Por otro, Dennett presenta experimentos del pensamiento (como el de Chase y Sanborn), así como casos reales (como los de acromatopsia cerebral), en los que la infalibilidad e incorregibilidad y, por ende, la autoridad de primera persona y el acceso directo en la conciencia, quedan fuertemente cuestionados. Parte del problema surge de que en ambos tipos de casos es imposible aislar los demás estados y procesos mentales de las experiencias, de suerte que uno pueda saber si cierta variación a la que uno está respondiendo es una variación en experiencia o en, digamos, memoria. Tomemos nuevamente el caso intrapersonal del espectro invertido. En este caso suponemos que S'''' ha sufrido una inversión de su espectro visual porque mediante la introspección —decimos— él se da cuenta de la inversión: las cosas ya no le parecen como solían parecerle. Pero esto sólo lo cree el sujeto con base en ciertos recuerdos que tiene sobre cómo le parecían las cosas. La pregunta que surge es si nuestra conjetura inicial es correcta, si la introspección de S'''' es infalible en este aspecto. ¿No se produciría la misma reacción en el sujeto si, en vez de alterarse su espectro, fuera su memoria la que se alterara? Seguramente que sí, y esto indica que la introspección no es el medio infalible para saber qué le ha ocurrido a S''''. Una investigación de su neurofisiología y otros estudios desde la perspectiva de la tercera persona ayudarían a resolver la cuestión. Sin embargo, una vez que hemos admitido la ayuda de la tercera persona para resolver qué le ha ocurrido a S'''', ya hemos dejado de lado, según Dennett, la idea de *qualia* como tradicionalmente se la concebía.

Quizá Dennett esté en lo correcto al dudar de la noción de *qualia*. Después de todo, presentan un reto formidable al fisicalismo y al funcionalismo en cuanto constituyen algo que no nos permite cerrar el hueco explicativo. No obstante, hay dos réplicas a esta estrategia argumentativa.

Primero, el que un concepto preteórico o teórico nos lleve a paradojas —sean o no formales— no es razón suficiente para deshacernos del concepto. El que el hueco explicativo persis-

ta, así como el que haya ciertas "paradojas", sólo presenta un reto al partidario de los *qualia*. Segundo, hay algo indudable de las experiencias y es que al momento de tenerlas sé cómo se sienten, aunque no *sepa* en todas las situaciones (esto se lo concedemos a Dennett) si son iguales o no a las que tenía hace unos días o incluso hace unos momentos. Dennett tiene que darnos, pues, algo que sustituya a los *qualia* sin conllevar los problemas que a ellos les atribuye y que dé cuenta de todas las experiencias. Propone entender los estados de experiencia como estados informacionales de propiedades o rasgos en el mundo. El que un estado o suceso lleve información acerca de una propiedad como un color en el mundo es algo *objetivo* cuanto es independiente de las creencias que el sujeto tiene sobre la información que dicha experiencia lleva. Una experiencia, además, puede causar en el sujeto actos de aparente, o en ciertos casos, real, reidentificación o discriminación de la propiedad. El poder de causar estos actos en el sujeto es una propiedad disposicional que le da a él "acceso" a las experiencias de las propiedades en el mundo. La propuesta de Dennett es, en suma, una propuesta intencional en cuanto informacional de la experiencia, y disposicional o funcional por las reacciones que los estados de experiencia causan.

Hay dos asuntos que señalar sobre la propuesta de Dennett. Primero, si ha de ser exitosa, será necesario mostrar que la manera en que nos percatamos de nuestras experiencias no constituye más que el que lleguemos a realizar actos (aparentes o reales) de reidentificación o discriminación.[30] Segundo, como propuesta informacional, puede ser que la de Dennett funcione para explicar las experiencias perceptuales. Es más, los ejemplos de *qualia* que él mismo presenta en su artículo son todos ellos ejemplos de propiedades de experiencias que se derivan de los sentidos. No hay ahí una discusión de otras experiencias no derivadas de los sentidos como pueden ser los

[30] Es una pregunta adicional e interesante en qué consisten estos actos de reidentificación o discriminación. ¿Consisten, acaso, en creencias que nos formamos? ¿Son las creencias cualitativas de las que habla Shoemaker? En caso afirmativo, la pregunta sería, entonces, si la conciencia fenoménica, esto es, cómo se sienten las experiencias, consiste simplemente en tener creencias sobre esas experiencias.

dolores u otras sensaciones, sensaciones que a primera vista, al menos, *no* tienen un contenido informacional. Queda abierta entonces la pregunta de si Dennett ha logrado darnos algo que pueda sustituir adecuadamente a los *qualia* que tanto repudia, por lo menos para el caso de este tipo de experiencias. El asunto dependerá, en parte, de si se puede atribuir contenido intencional a sensaciones como los dolores y los cosquilleos.

Como veremos en el caso de las propuestas también intencionales de Harman y de Tye sobre el carácter cualitativo de la experiencia, es un reto para este tipo de sugerencias explicar en qué sentido los dolores y cosquilleos son estados intencionales.

5. *Más sobre el funcionalismo e inversiones*

Gilbert Harman, al igual que Dennett, niega que los rasgos cualitativos sean accesibles a la conciencia. En su artículo ("La cualidad intrínseca de la experiencia") se pregunta si el funcionalismo puede dar cuenta de la sensación subjetiva de la experiencia, del "cómo es" tener tal o cual experiencia. Para ello considera tres argumentos relacionados que pretenden demostrar que el funcionalismo no puede dar cuenta de este aspecto de la experiencia, y arguye que los tres argumentos son falaces.

En el primer argumento se afirma que el funcionalismo no puede dar cuenta de la cualidad intrínseca de la experiencia, ya que le conciernen solamente las relaciones. La respuesta de Harman es que falla porque confunde las características intrínsecas *del objeto intencional* de la experiencia con las características intrínsecas *de la experiencia misma*. Pero uno no es consciente del carácter intrínseco de su experiencia y, por eso, el hecho de que el funcionalismo se abstraiga del carácter intrínseco de la experiencia no demuestra que deje fuera algo de lo que uno es consciente. La aseveración de que normalmente somos conscientes de las propiedades intrínsecas de nuestras experiencias involucra una falacia que se origina en una confusión entre las propiedades del objeto representado y las propiedades de una representación de ese objeto. De acuerdo con Harman, al percibir un objeto somos conscientes de las propiedades del objeto de nuestra experiencia, esto es, del contenido de la representación, pero no de las propiedades intrínsecas de nuestra

experiencia, esto es, no de las propiedades del vehículo de la representación.

El segundo argumento es una forma del argumento del conocimiento: una persona ciega de nacimiento podría saber acerca de los hechos físicos y funcionales de la percepción del color, sin saber cómo es ver algo rojo. Por lo tanto, la experiencia de ver rojo no puede explicarse en términos puramente funcionales. La defensa de Harman frente a este argumento difiere de la de Lewis en que acepta que alguien que tiene experiencias del color rojo que antes no tenía adquiere nuevo conocimiento. Pero también rechaza el supuesto de Jackson de que María podía saber todo lo físicamente pertinente a la visión antes de salir de la habitación. Harman argumenta que una persona que ha vivido siempre en un entorno sin color rojo, simplemente no podría haber adquirido el concepto de rojo, dado que tener este concepto significa estar funcionalmente determinado de cierta manera y esto sólo puede darse si uno ha sido expuesto, perceptualmente, a objetos rojos. Para que el concepto funcione de manera tal que la persona tenga el concepto pleno de rojo, ésta debe estar dispuesta a formarse representaciones que involucren ese concepto como la consecuencia natural e inmediata de ver algo rojo. Como la persona ciega, o María, carece de dicha experiencia perceptual, no podría haber adquirido ese concepto, y por lo tanto, no tiene *todo* el conocimiento pertinente acerca de las experiencias de los colores. Conocer cómo es ver algo rojo es ser capaz de representar algo rojo, y las representaciones mentales se construyen a partir de conceptos; y el contenido de una representación se determina por los conceptos que contiene y la forma en que éstos se integran para formar esa representación.

La respuesta de Harman a la segunda objeción sólo es válida para un tipo de funcionalismo que hace referencia a las funciones de los conceptos y no sólo a las funciones de los estados mentales en general. Así, como admite Harman, alguien como Lewis se ve obligado a decir que, aunque un ciego de nacimiento carece de ciertas habilidades, no carece de ninguna información sobre la percepción visual, cosa que a Harman le parece falsa.

El tercer argumento es el que afirma que el funcionalismo no puede dar cuenta de la posibilidad de un espectro invertido. Harman considera este argumento tanto en su versión interpersonal como en su versión intrapersonal (véanse los artículos de Shoemaker "Funcionalismo y *qualia*" y Shoemaker 1982). Sostiene que el funcionalismo puede rechazar la posibilidad de un espectro invertido si rechaza la tesis de que la percepción requiere estar consciente de ciertas características *intrínsecas* de la experiencia. Su defensa frente a la primera objeción justifica el rechazo a dicha tesis. Para él, en la percepción normal no hay una distinción entre las propiedades intrínsecas y las intencionales de una experiencia. Pensar que se pudieran presentar casos pertinentes de espectro invertido sin una inversión de las creencias y los significados es caer otra vez en la primera objeción. Así, la posibilidad de un espectro invertido requiere que haya una inversión de creencias y significados y, dado que esto último puede capturarse en un marco funcionalista, el argumento del espectro invertido no presenta dificultad alguna al funcionalismo.

En respuesta a los argumentos de Harman, Block ("La Tierra Invertida") imagina un nuevo tipo de inversión que pretende mostrar que los *qualia* son rasgos mentales intrínsecos de nuestra experiencia y que, por tanto, no son funcionalmente caracterizables. Pero, a pesar de ello, Block sigue siendo un cuasifuncionalista porque considera que el contenido intencional de la experiencia es funcional. Admite la posibilidad de que dos experiencias sean funcionalmente idénticas pero con distinto contenido cualitativo (*qualia* invertidos), o que difieran funcionalmente y tengan el mismo contenido cualitativo. El caso de la Tierra Invertida es un ejemplo de esto último. Muestra que el argumento del conocimiento (que María puede saber todo lo físicamente pertinente acerca de la visión sin tener experiencias de color) no sirve para decidir entre aquellos que, como Harman, creen que los *qualia* de una experiencia simplemente son sus propiedades intencionales y aquellos que lo niegan. Block piensa que el caso de la Tierra Invertida sí sirve para decidir entre estas dos posiciones. El problema con Harman es que comete lo que Block llama la falacia de la *intencionalización*

de los qualia: identifica las propiedades intencionales con las cualitativas.

El experimento de pensamiento de la Tierra Invertida se construye para mostrar que puede haber cambios en el contenido intencional sin cambios en los *qualia* y reivindica una de las características de las experiencias que Dennett y Harman creían rechazar, a saber, las cualidades intrínsecas y mentales que son parte de la caracterización de los *qualia*. El experimento es el siguiente: imaginemos que, sin advertirlo, nos ponen unos lentes que invierten los colores, al tiempo que nos trasladan a la Tierra Invertida. La Tierra Invertida difiere de la tierra en que los colores de los objetos, así como el vocabulario sobre los colores, están invertidos con respecto a los de la tierra. Frente a las de los terrícolas, las experiencias de los habitantes de la tierra invertida tienen invertido tanto el contenido cualitativo como el contenido intencional. Block sostiene que en la Tierra Invertida, nuestras experiencias (con los lentes inversores) serían cualitativamente iguales a las que teníamos en la tierra y también, al principio, nuestros contenidos intencionales. Sin embargo, después de un periodo de adaptación a nuestro nuevo entorno físico y lingüístico en la Tierra Invertida, nuestros contenidos intencionales y funcionales cambiarían hasta llegar a ser los mismos que los de los nativos.[31] Tendríamos, pues, un caso de mismos *qualia* pero distinto contenido intencional. Frente a la objeción de Harman contra el espectro invertido, el caso de la Tierra Invertida reivindica la posibilidad de separar las propiedades intencionales de los *qualia*, y con ello le da nueva vida al argumento original del espectro invertido contra el funcionalismo.

Para resumir, podemos afirmar que ante las dificultades planteadas por la posibilidad de *qualia* invertidos y de *qualia* ausentes, se perfilan tres grupos de posiciones que hemos discutido aquí. La primera sería la posición de Block, la cual sostiene que si las situaciones que se describen en los experimentos antes mencionados son plausibles, entonces el "cómo es ser" un or-

[31] Aquí se está suponiendo una suerte de externalismo de los contenidos intencionales, esto es, que dichos contenidos se individúan en términos de los entornos físicos externos en los que se encuentran los sujetos.

ganismo que tiene experiencias no puede ser captado por los análisis funcionalistas. El segundo grupo está representado por Harman y Dennett y es más radical en el sentido de que reduce o niega la existencia de los *qualia*, aunque ellos dan diversas versiones de por qué *parece* que los tenemos. En tercer lugar, están, por un lado, las posturas compatibilistas que quieren dar cuenta de los *qualia* mediante análisis funcionalistas y, por otro, los que insisten en que los *qualia* se deben, en parte, a las estructuras físicas que los realizan y, por tanto, no son asunto que el funcionalismo tenga que explicar. El ejemplo más claro del primer tipo de tesis es el de Shoemaker, que ya hemos caracterizado. Los que adoptan el segundo tipo de compatibilismo, son, por ejemplo, David Lewis (1980) y Paul Churchland (1984), quienes sostienen que los rasgos cualitativos intrínsecos se identifican con propiedades neurofisiológicas. Su función sería facilitar una identificación introspectiva rápida de las sensaciones, pero éstas no serían esenciales para determinar los tipos de estados mentales que deberán determinarse funcionalmente. El problema con una sugerencia como la de Lewis y Churchland es que si identificamos los *qualia* con estados neurofisiológicos que desempeñan los papeles funcionales adecuados, volvemos a las dificultades que ya hemos mencionado en relación con el fisicalismo, a saber, que simplemente no está claro cómo una explicación fisicalista (esto es, en términos neurofisiológicos o de cualquier otra ciencia física) pueda darnos una explicación composicional o constitutiva de por qué un estado de dolor se siente como se siente.

6. *Una propuesta representacionista, los retos y las perspectivas*

Hemos visto ya que Harman sugiere que las experiencias no tienen cualidades intrínsecas sino sólo intencionales. Al parecer, esta propuesta intencionalista tiene cierta plausibilidad cuando la ejemplificación de un tipo de sensación puede constituir una percepción.[32] Éste era el caso de las sensaciones del rojo 13 o del rojo 14. La propuesta parece más cuestionable cuando se aplica a sensaciones que al ser instanciadas, no parecen consti-

[32] Este tema se verá con mayor detalle en el volumen II sobre percepción.

tuir percepciones. Por ejemplo, las sensaciones de dolor o de cosquilleo. ¿Cuál sería el contenido de dichas sensaciones?

En una línea intencionalista, o más claramente representacionista, Michael Tye ("Una teoría representacional del dolor y de su carácter fenoménico") pretende dar una versión del contenido de las sensaciones (sean percepciones o no) que se identifican con los *qualia*. Sugiere que todas las experiencias tienen siempre un contenido representacional preconceptual o no conceptual, y que dicho contenido agota la conciencia fenoménica. Sus argumentos más interesantes son en relación con el fenómeno del dolor, ya que argumenta, en contra de las opiniones de la mayoría de los filósofos, que el dolor sí es objeto de estudio para la psicología cognoscitiva. Lo es, porque los dolores tienen un contenido representacional, y su carácter fenoménico se utiliza en la operación de sistemas cognoscitivos (*i.e.*, sistemas que operan con base en una psicología de creencias y deseos).

Para Tye, los dolores son estados intencionales, esto es, representaciones cuyo contenido intencional es no conceptual.[33] Como estados con contenido intencional son capaces de representar erróneamente. Esto se muestra, por ejemplo, en la manera en que hablamos de dolores en miembros amputados. Admitimos como verdadera la afirmación de que las personas están sujetas a dolores en miembros que ya no existen. El contenido intencional del dolor, según Tye, concierne a la pierna, la cual puede o no existir. Así, uno puede sentir dolor en la pierna izquierda aun cuando ya no tenga pierna izquierda, tal y como se puede buscar la Fuente de la Juventud sin que ésta exista. Las experiencias del dolor, si acaso se encuentran en algún lado, no es en las piernas o en los brazos, sino en el cerebro. Cuando decimos que tenemos un dolor en la pierna, lo que sucede es que la experiencia del dolor indica, cuando todo está funcionando normalmente, que hay un daño en la pierna. En

[33] El contenido representacional de los dolores puede ser bastante específico, dependiendo de las clases de dolores. Así, por ejemplo, una punzada es un dolor que representa un trastorno breve y ligero. Un dolor palpitante es el que representa un trastorno de pulsaciones rápidas. Un pinchazo... En cada uno de estos casos, el sujeto que siente el dolor experimenta una representación sensorial de cierta clase de trastorno corporal.

circunstancias normales, un dolor en la pierna es una experiencia sensorial concreta que representa correctamente que algo en la pierna está dañado.[34] Y uno se puede equivocar: puede sentir un dolor en el brazo izquierdo cuando no ocurre nada malo *en el brazo izquierdo*, sino que el trastorno tiene lugar en el corazón.

Pero, ¿cómo ha de explicarse lo doloroso de los dolores dentro de esta propuesta? A menudo hablamos de las lesiones corporales como dolorosas. Cuando se dice que un corte en el dedo o una quemadura son dolorosos o duelen, lo que quiere decirse, según Tye, es que están *causando* una sensación y que tal sensación provoca un desagrado inmediato junto con la preocupación por el estado de la parte del cuerpo donde la persona siente que se localiza la lesión. Cuando examinamos introspectivamente nuestros dolores, tomamos conciencia de sus contenidos sensoriales como dolorosos. Mi dolor representa una lesión en mi pierna, y clasifico entonces cognoscitivamente esa lesión como dolorosa, aplicando el concepto de *doloroso* en la introspección.

Tye da respuesta a la pregunta de qué clases de representaciones son los dolores. Ya hemos dicho que para él son representaciones no conceptuales, pero entonces, ¿de qué manera representan entonces? Dado que la corteza somatosensorial está organizada topográficamente y es la localización primaria del dolor, Tye propone un modelo de los dolores como representaciones topográficas trazadas como mapas. O, en términos más específicos, sostiene que "los dolores son patrones de células activas que tienen lugar en ordenamientos tridimensionales, topográficamente estructurados, a los que se anexan etiquetas descriptivas" (p. 340).

Para el caso de cualquier sensación corporal, como la experiencia perceptual de una forma cuadrada y/o azul, o de un dolor, se requiere en categorizaciones como representaciones de uno u otro, pero éstas son, según Tye, automáticas y no exigen que la criatura tenga creencias o pensamientos sobre las

[34] Tye identifica los dolores con sucesos que son experiencias concretas. Los sucesos se individúan en parte a través de los objetos que los padecen y esto explica por qué ninguna otra persona podría sentir el dolor que estoy sintiendo y por qué los dolores los tiene una persona particular.

propiedades de los estímulos. De este modo, no hay nada en la explicación que implique que un niño muy pequeño o un animal no puedan sentir dolor. No es necesario ningún pensamiento (o creencia) acerca de las características reales o aparentes del objeto observado. Esto no quiere decir, sin embargo, que la experiencia del dolor no pueda penetrar cognoscitivamente. Los contenidos están ligados a las representaciones de salida del o de los módulos sensoriales pertinentes y se encuentran listos y en posición de ejercer un impacto directo en el sistema de creencias y deseos.

La propuesta representacionista de Tye es, no cabe duda, la propuesta más detallada de las propuestas intencionalistas. No sólo da una explicación del tipo de representación mental que son las experiencias, basada en teorías empíricas de las ciencias cognoscitivas actuales, sino que –a diferencia de Harman– se enfrenta directamente al reto de explicar cuál sería el contenido representacional de sensaciones que a primera vista no tienen contenido intencional. Surge la pregunta entonces de si hay un hueco explicativo aquí.

A lo largo de esta introducción hemos recorrido varios de los temas y problemas que consideramos más pertinentes en la discusión actual sobre las sensaciones y, particularmente, sobre su aspecto fenoménico. Hemos visto que hay un hueco explicativo entre lo físico y lo mental, y entre lo funcional y lo mental, que ha de cerrarse si es que tenemos alguna esperanza de poder explicar las sensaciones composicional y/o causalmente. Éste constituye uno de los retos más importantes para la filosofía de la mente, y se exhibe de distintos modos. Se exhibe como el hueco entre lo subjetivo y lo objetivo, y como el hueco entre lo físico y lo funcional, por un lado, y lo fenoménico, por el otro. Estas dos maneras de caracterizar el hueco explicativo son distintas.

La primera concierne al uso de conceptos objetivos para dar una explicación de algo esencialmente subjetivo; y entre los conceptos objetivos podrían incluirse los físicos o los funcionales, pero también otros que fueran independientes de la perspectiva que tiene un sujeto sobre sus propias experiencias. Podríamos quizá intentar caracterizar las experiencias de alguien usando conceptos intencionales como lo sugiere Dennett, o incluso

predicados de creencia, los cuales (supongamos por mor del argumento) no dependieran esencialmente del punto de vista del sujeto para ser comprendidos. En estos casos diríamos (aunque en un sentido limitado) que los términos son objetivos, y el hueco explicativo persistiría ya que no lograríamos dar una explicación plena de aquello que depende esencialmente del punto de vista del sujeto. No así en la segunda manera de caracterizar el hueco. Ésta no supone que el hueco explicativo se deba a los diferentes puntos de vista subjetivos, sino a que hay ciertas características de lo físico y/o de lo funcional, exhibidas en los argumentos del conocimiento y en los experimentos del pensamiento (las diversas inversiones), que no les permiten explicar constitutivamente la conciencia fenoménica. Estos argumentos intentan mostrar no tanto que no se podrá cerrar el hueco explicativo porque los conceptos necesarios para el *explanandum* requerirían una perspectiva subjetiva mientras que los del *explanans* no la requieren, sino que por muy detalladas que fueran las explicaciones neurofisiológicas o las funcionales no nos podrán decir cómo a partir de algo físico o funcional surge determinado suceso o estado con fenomenología.

Las propuestas más recientes sobre los *qualia* son las propuestas intencionalistas como las de Dennett, Harman y Tye. A diferencia de Harman y Tye, Dennett piensa que no hay algo como el carácter cualitativo de las sensaciones, y que a lo sumo tenemos ciertas reacciones a las experiencias con cierto contenido informacional. Harman y Tye, en cambio, piensan que sí hay un contenido cualitativo, pero que éste no difiere del contenido intencional. El problema es cómo dar cuenta de las experiencias que al parecer carecen de un contenido intencional. Tye ofrece una respuesta: sensaciones, como los dolores, simplemente representan que algo desagradable está sucediendo en el cuerpo, que un daño ha ocurrido en una parte del cuerpo. La pregunta que surge es cómo debemos interpretar este contenido intencional. Recordemos que una de las características del carácter fenoménico de las experiencias es que hay cierto acceso directo a él, que un sujeto sabe, en un momento dado (concediendo que este conocimiento pueda ser sólo momentáneo para evitar las preocupaciones de Dennett), cuál es el carácter fenoménico de su experiencia, esto es, cómo es su

experiencia. Ahora bien, si el carácter fenoménico de los dolores se reduce al contenido (no conceptual) de que hay cierto daño en el cuerpo, deberíamos poder conocer esto por vía de la introspección. Ciertamente, sé cómo se siente mi dolor, y sé también que es un indicio de que algo no anda bien en mi cuerpo. Pero, ¿son lo mismo? Y más aún, ¿se puede identificar un contenido representacional semejante para otras sensaciones como cosquilleos o ardores? O, ¿persiste un hueco explicativo entre lo intencional y lo cualitativo? Éstas, entre muchas otras preguntas, quedan abiertas.

BIBLIOGRAFÍA[36]

Burge, T., 1997, "Two Kinds of Consciousness", en Block, Flanagan y Güzeldere 1997, pp. 427-433.*

Block, N., 1995, "On a Confusion about a Function of Consciousness", *Behavioral and Brain Sciences*, vol. 18, no. 2, pp. 227-287.*

———, 1980a, "Are Absent Qualia Impossible?", *Philosophical Review*, vol. 89, pp. 257-274.

———, 1980b, *Readings in the Philosophy of Psychology*, vol. I, Harvard University Press, Cambridge, Mass.

———, 1980c, "Troubles with Functionalism", en Block 1980b, pp. 268-305.*

———, O. Flanagan y G. Güzeldere (comps.), 1997, *The Nature of Consciousness. Philosophical Debates*, MIT Press, Cambridge, Mass.*

Chalmers, D., 1995, "Absent Qualia, Fading Qualia, Dancing Qualia", en Metzinger 1995, pp. 309-328.

Churchland, P., 1989a, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Mass.

———, 1989b, "Knowing Qualia: A Reply to Jackson", en Churchland 1989a, pp. 67-76.*

———, 1985, "Reduction, Qualia and the Direct Introspection of Brain States", *Journal of Philosophy*, vol. 82, no. 1, pp. 8-28.

Churchland, P., 1984, *Matter and Consciousness*, MIT Press, Cambridge, Mass. [Versión en castellano: *Materia y conciencia*, trad. M.N. Mizraji, Gedisa, Barcelona, 1992.]

[36] Las referencias bibliográficas con asterisco se deberán considerar también como lecturas adicionales recomendadas.

Conee, E., 1985, "The Possibility of Absent Qualia", *Philosophical Review*, vol. 94, pp. 345–366.*

Crane, T., 1991, "Why Indeed? Papineau on Supervenience", *Analysis*, vol. 51, no. 1, pp. 32–37.

Davidson, D., 1970, "Mental Events", en Foster y Swanson 1970, pp. 79–101. [Versión en castellano: "Sucesos mentales", en *Ensayos sobre acciones y sucesos*, trad. O. Hansberg, J.A. Robles y M.M. Valdés, Crítica/Instituto de Investigaciones Filosóficas-UNAM, Barcelona, 1995, pp. 263–287.]

——— y G. Harman (comps.), 1972, *Semantics of Natural Language*, Reidel, Dordrecht.

Davies, M. y G. Humphreys (comps.), 1993, *Consciousness*, Blackwell, Oxford.*

Descartes, R., 1637 (fecha de publicación), *Meditaciones metafísicas y otros*, trad. E. López y M. Grana, Gredos, Madrid, 1987.

Fodor, J., 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge, Mass. [Versión en castellano: *Psicosemántica: El problema del significado en la filosofía de la mente*, trad. Ó.L. González-Castan, Tecnos, Madrid, 1994.]

Foster, L. y J. Swanson (comps.), 1970, *Experience and Theory*, Duckworth, Londres.

Kim, J., 1993, *Supervenience and the Mind: Selected Philosophical Essays*, Cambridge University Press, Cambridge.

———, 1984, "Concepts of Supervenience", *Philosophy and Phenomenological Research*, vol. 45, no. 2, pp. 153–176. (También publicado en Kim 1993, pp. 53–78.)

Kripke, S., 1972, "Naming and Necessity", en Davidson y Harman 1972, pp. 253–355. (Versión en libro: *Naming and Necessity*, Harvard University Press, Cambridge, Mass., 1980. [Versión en castellano: *El nombrar y la necesidad*, 2a. ed., trad. M.M. Valdés, Instituto de Investigaciones Filosóficas-UNAM, 1995.]*

Leibniz, G.W., 1765 (fecha de publicación, pero escrito en 1704), *Nuevos ensayos sobre el entendimiento humano*, trad. J. Echeverría Ezponda, Nacional, Madrid, 1983.

Levine, J., 1988, "Absent and Inverted Qualia Revisited", *Mind and Language*, vol. 3, pp. 271–287.*

———, 1983, "Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly*, vol. 64, pp. 354–361.

Lewis, D., 1980, "Mad Pain and Martian Pain", en Block 1980b, pp. 216–222.

Loar, B., 1990, "Phenomenal States", *Philosophical Perspectives*, vol. 4, pp. 81–108.*

Metzinger, T. (comp.), 1995, *Conscious Experience*, Ferdinand Schoningh, Paderborn.*
Papineau, D., 1991, "The Reason Why: Response to Crane", *Analysis*, vol. 51, no. 1, pp. 37-40.
———, 1990, "Why Supervenience?", *Analysis*, vol. 50, no. 2, pp. 66-71.
Rosenthal, D., 1986, "Two Concepts of Consciousness", *Philosophical Studies*, vol. 49, pp. 329-359.
Sexto Empírico, *Esbozos pirrónicos*, trad. A. Gallego Cao y T. Muñoz Diego, Gredos, Madrid, 1993.
Shoemaker, S., 1984, *Identity, Cause, and Mind*, Cambridge University Press, Cambridge.
———, 1982, "The Inverted Spectrum", *Journal of Philosophy*, 327-357.*
———, 1981, "Absent Qualia Are Impossible —A Reply to Block", *Philosophical Review*, vol. 90, pp. 581-600. (También en Shoemaker 1984, pp. 309-326.)*
Tye, M., 1986, "The Subjective Qualities of Experience", *Mind*, vol. 95, no. 377, pp. 1-17.*
Van Gulick, R., 1993, "Understanding the Phenomenal Mind: Are We All Just Armadillos?", en Davies y Humphreys 1993.

LECTURAS ADICIONALES RECOMENDADAS

Akins, K., 1993, "What Is It Like to Be Boring and Myopic?", en Dahlbom 1993.
Biro, J.I., 1993, "Consciousness and Objectivity", en Davies y Humphreys 1993.
Chalmers, D., 1996, *The Conscious Mind*, Oxford University Press, Oxford. [Versión en castellano: *La mente consciente*, trad. J.A. Álvarez, Gedisa, Barcelona, 1999.]
Dahlbom, B. (comp.), 1993, *Dennett and his Critics*, Blackwell, Oxford.
Dennett, D., 1991, *Consciousness Explained*, Little Brown, Boston. [Versión en castellano: *La conciencia explicada*, trad. S. Balari Ravera, Paidós, Barcelona, 1995.]
Dretske, F., 1993, "Conscious Experience", *Mind*, vol. 102, no. 406, pp. 263-283.
Flanagan, O., 1992, *Consciousness Reconsidered*, MIT Press, Cambridge, Mass.
Garvey, J., 1997, "What Does McGinn Think We Cannot Know?", *Analysis*, vol. 57, no. 3, pp. 196-201.
Hanson, P.P., 1993, "McGinn's Cognitive Closure", *Dialogue*, vol. 32, no. 3, pp. 579-585.

Hookway, C. y D. Peterson (comps.), 1993, *Philosophy and Cognitive Science*, Cambridge University Press, Cambridge (Royal Institute of Philosophy Supplement, 34).

Horgan, T., 1987, "Supervenient Qualia", *Philosophical Review*, vol. 96, pp. 491–520.

———, 1984, "Jackson on Physical Information and Qualia", *Philosophical Quarterly*, vol. 34, pp. 147–151.

Levine, J., 1997a, "Are Qualia Just Representations?", *Mind and Language*, vol. 12, no. 1, pp. 101–113.

———, 1997b, "Recent Work on Consciousness", *American Philosophical Quarterly*, vol. 34, no. 4, pp. 379–404.

———, 1995, "Qualia: Intrinsic, Relational, or What?", en Metzinger 1995, pp. 277–292.

———, 1993, "On Leaving Out What It Is Like", en Davies y Humphreys 1993.

Levin, J., 1991, "Analytic Functionalism and the Reduction of Phenomenal States", *Philosophical Studies*, vol. 61, pp. 211–238.

Lormand, E., 1994, "Qualia! (Now Showing at a Theater Near You)", *Philosophical Topics*, vol. 22, pp. 127–156.

Lycan, W., 1996, *Consciousness and Experience*, MIT Press, Cambridge, Mass.

———, 1987, *Consciousness*, MIT Press, Cambridge, Mass.

Marcel, A., 1988, "Phenomenal Experience and Functionalism", en Marcel y Bisiach 1988, pp. 121–158.

———y E. Bisiach (comps.), 1988, *Consciousness in Contemporary Science*, Oxford University Press, Oxford.

McGinn, C., 1991, *The Problem of Consciousness: Essays Toward a Resolution*, Blackwell, Oxford.

Mellor, D.H., 1993, "Nothing Like Experience", *Proceedings of the Aristotelian Society*, vol. 93, pp. 1–16.

Neander, K., 1998, "The Division of Phenomenal Labor: A Problem for Representationalist Theories of Consciousness", *Philosophical Perspectives*, vol. 12, pp. 411–434.

Rey, G., 1998, "A Narrow Representationalist Account of Qualitative Experience", *Philosophical Perspectives*, vol. 12, pp. 435–458.

Robinson, D., 1993, "Epiphenomenalism, Laws, and Properties", *Philosophical Studies*, vol. 69, no. 1, pp. 1–34.

Searle, J., 1992, *The Rediscovery of the Mind*, MIT Press, Cambridge, Mass. [Versión en castellano: *El redescubrimiento de la mente*, trad. L.M. Valdés Villanueva, Crítica, Barcelona, 1996.]

Shoemaker, S., 1993, "Lovely and Suspect Ideas", *Philosophy and Phenomenological Research*, vol. 53, no. 4, pp. 905–910.

Shoemaker, S., 1991, "Qualia and Consciousness", *Mind*, vol. 100, no. 4, pp. 507–524.

———, 1990, "Qualities and Qualia: What's in the Mind", *Philosophy and Phenomenological Research*, vol. 50, suplemento, pp. 109–131.

Stoljar, D., 1996, "What What It's Like Isn't Like", *Analysis*, vol. 56, no. 4, pp. 281–283.

Tye, M., 1999, "Phenomenal Consciousness: The Explanatory Gap As a Cognitive Illusion", *Mind*, vol. 108, no. 432, pp. 705–725.

———, 1998, "Inverted Earth, Swampman, and Representationalism", *Philosophical Perspectives*, vol. 12, pp. 459–477.

———, 1995, *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, MIT Press, Cambridge, Mass.

———, 1993a, "Blindsight, the Absent Qualia Hypothesis, and the Mystery of Consciousness", en Hookway y Peterson 1993, pp. 19–40.

———, 1993b, "Reflections on Dennett and Consciousness", *Philosophy and Phenomenological Research*, vol. 53, no. 4, pp. 893–898.

¿CÓMO ES SER UN MURCIÉLAGO?*

THOMAS NAGEL

La conciencia es lo que vuelve el problema mente-cuerpo realmente inextricable. Quizá por ello las discusiones actuales del problema le prestan poca atención o la entienden de una manera claramente equivocada. La reciente ola de euforia reduccionista ha producido varios análisis de los fenómenos y conceptos mentales destinados a explicar la posibilidad de alguna variante del materialismo, de la identificación psicofísica o de la reducción.[1] Pero los problemas que se abordan son los comunes a esta y otras clases de reducción, y se desatiende lo que hace único al problema mente-cuerpo y lo distingue del problema agua-H_2O, el problema máquina de Turing-máquina IBM, el problema rayo-descarga de electricidad, el problema genes-ADN o el problema roble-hidrocarbón.

Cada reduccionista tiene su analogía favorita que toma de la ciencia moderna. Es sumamente improbable que alguno de estos ejemplos inconexos de reducción exitosa arroje luz sobre la relación de la mente con el cerebro. Pero los filósofos comparten la debilidad humana general de formular explicaciones

* He leído distintas versiones de este artículo ante varios auditorios, y estoy en deuda con muchas personas por sus comentarios.

[1] Algunos ejemplos son: Smart 1963; Lewis 1966; Putnam 1967; Armstrong 1968; Dennett 1969. He formulado dudas respecto a este tipo de posiciones anteriormente en Nagel 1970; Nagel 1971 y una reseña de Dennett, *Journal of Philosophy*, 69, 1972. Véase, además, Kripke 1972, esp. las pp. 334–342, y Thornton 1972.

de lo que es incomprensible en términos adecuados para lo que es conocido y bien comprendido, aunque se trate de algo totalmente distinto. Esto ha llevado a la aceptación de explicaciones poco verosímiles de lo mental, en gran medida porque permitirían formas conocidas de reducción. Trataré de explicar por qué los ejemplos usuales no nos ayudan a comprender la relación entre la mente y el cuerpo; por qué, de hecho, no tenemos actualmente una concepción de lo que sería una explicación de la naturaleza física de un fenómeno mental. Sin la conciencia, el problema mente-cuerpo sería mucho menos interesante; con la conciencia parece no tener solución. Comprendemos de una manera muy pobre el rasgo más importante y característico de los fenómenos mentales conscientes. La mayoría de las teorías reduccionistas ni siquiera intentan explicarlo, y se puede mostrar con un examen cuidadoso que ningún concepto de reducción disponible hoy día puede aplicársele. Quizá se pueda concebir una forma teórica nueva para ese fin; pero tal solución, si acaso existe, se encuentra en un futuro intelectual distante.

La experiencia consciente es un fenómeno generalizado. Se encuentra en muchos niveles de la vida animal, aunque no podamos estar seguros de su presencia en los organismos más simples, y es muy difícil afirmar en general qué constituiría un indicio de su existencia. (Algunos extremistas están dispuestos a negar su existencia incluso en los mamíferos no humanos.) Sin duda se presenta en numerosas formas totalmente inimaginables para nosotros, en otros planetas de otros sistemas solares a lo largo del universo. Pero al margen de cómo varíe la forma, el hecho de que un organismo *tenga* experiencias conscientes significa, básicamente, que hay algo que es cómo es *ser* ese organismo. Puede haber más implicaciones sobre la forma de la experiencia; incluso puede haber (aunque lo dudo) implicaciones sobre la conducta del organismo. Pero, fundamentalmente, un organismo tiene estados mentales conscientes si y sólo si hay algo que es cómo es *ser* ese organismo, algo que es cómo es ser *para* ese organismo.

Podemos llamar a esto el carácter subjetivo de la experiencia. Se trata de algo que no es captado por ninguno de los análisis reduccionistas recientes más conocidos de los fenómenos men-

tales, porque todos ellos son lógicamente compatibles con su ausencia. No puede analizarse en términos de ningún sistema explicativo de estados funcionales o de estados intencionales, pues éstos podrían atribuirse a robots o a autómatas que se comportaran como personas, aunque no tuvieran ningún tipo de experiencia.[2] Debido a razones similares, tampoco puede analizarse en términos del papel causal de las experiencias en relación con la conducta humana típica.[3] No niego que los estados y los sucesos mentales conscientes causen la conducta, ni que se les pueda dar una caracterización funcional. Sólo niego que ello agote su análisis. Todo programa reduccionista tiene que basarse en un análisis de lo que debe ser reducido. Si el análisis deja fuera algo, el problema se planteará de manera equivocada. Es inútil basar la defensa del materialismo en un análisis de los fenómenos mentales que no aborde de una manera explícita su carácter subjetivo. Pues no hay ninguna razón para suponer que una reducción que nos parece plausible cuando no se ha intentado explicar la conciencia pueda extenderse para que incluya a la conciencia. Por lo tanto, si no tenemos una idea de qué es el carácter subjetivo de la experiencia, no podemos saber qué se le exige a la teoría fisicalista.

Si bien una explicación de la base física de la mente debe explicar muchas cosas, el carácter subjetivo de la conciencia parece ser la más difícil. Es imposible excluir de una reducción las características fenomenológicas de la experiencia del mismo modo en que excluimos las características fenoménicas de una sustancia ordinaria en una reducción química o física de ella; es decir, explicándolas como efectos en la mente de observadores humanos.[4] Si hay que defender el fisicalismo, debemos ofrecer una explicación física de las características fenomenológicas.

[2] Quizá en realidad no podrían existir tales robots. Quizá cualquier cosa suficientemente compleja como para comportarse como una persona tendría experiencias. Pero ello, si es cierto, es un hecho que no puede ser descubierto mediante el simple análisis del concepto de experiencia.

[3] No es equivalente a aquello en relación con lo cual somos incorregibles, porque no somos incorregibles con respecto a la experiencia y porque la experiencia es algo que está presente en animales que carecen de lenguaje y pensamiento, que no poseen creencia alguna sobre sus experiencias.

[4] *Cfr.* Rorty 1965, en especial las pp. 37–38.

Pero cuando examinamos su carácter subjetivo, tal resultado parece imposible de alcanzar. La razón es que todo fenómeno subjetivo está conectado esencialmente con un solo punto de vista, y parece inevitable que una teoría física objetiva abandone ese punto de vista.

Permítaseme en primer lugar tratar de formular el problema de una manera más completa que simplemente refiriéndome a la relación entre lo subjetivo y lo objetivo, o entre el *pour soi* y el *en soi*. Esto dista de ser fácil. Los hechos acerca de lo que sería ser un *X* son muy peculiares, tan peculiares que algunos pueden verse inclinados a dudar de su realidad o de la importancia de las afirmaciones sobre ellos. Para ejemplificar la relación entre la subjetividad y un punto de vista, y para poner de manifiesto la importancia de los rasgos subjetivos, nos será útil explorar el asunto en relación con un ejemplo que haga resaltar de manera clara la divergencia entre los dos tipos de concepción: el subjetivo y el objetivo.

Supongo que todos creemos que los murciélagos tienen experiencias. Después de todo, son mamíferos, y no dudamos que tengan experiencias, así como tampoco dudamos que los ratones, las palomas y las ballenas las tengan. He elegido a los murciélagos y no a las avispas o los lenguados porque, si nos desplazamos a un nivel demasiado bajo en el árbol filogenético, las personas dejan de creer gradualmente que haya experiencias. Aunque están más relacionados con nosotros que esas especies, los murciélagos tienen un rango de actividad y un aparato sensorial tan diferente del nuestro que el problema que deseo plantear resulta muy vívido (aunque ciertamente podría plantearse con otras especies). Aun sin las ventajas de la reflexión filosófica, cualquiera que haya permanecido algún tiempo en un espacio cerrado con un murciélago alborotado sabe lo que significa encontrarse con una forma de vida fundamentalmente *ajena*.

He dicho que lo esencial de la creencia de que los murciélagos tienen experiencias es que hay algo que es cómo es ser murciélago. Hoy día sabemos que la mayoría de los murciélagos (los microquirópteros, para ser exactos) perciben el mundo externo básicamente mediante un sonar, o mediante

la ecolocalización, detectando los reflejos de sus propios chillidos rápidos, sutilmente modulados y de alta frecuencia, reflejos que provienen de los objetos a su alcance. Sus cerebros están diseñados para correlacionar los impulsos emitidos con los ecos subsecuentes, y la información así adquirida les permite hacer determinaciones precisas de distancia, tamaño, forma, movimiento y textura, comparables a las que hacemos con nuestra vista. Pero el sonar del murciélago, aunque es claramente una forma de percepción, no opera de manera similar a ninguno de nuestros sentidos, y no hay ninguna razón para suponer que se parece, en términos subjetivos, a algo que podamos sentir o imaginar. Esto aparentemente crea dificultades para la noción de cómo es ser un murciélago. Debemos considerar si hay algún método que nos permita proyectarnos hacia la vida interior del murciélago desde nuestro propio caso[5] y, si no lo hay, qué otros métodos podría haber para comprender esta noción.

Nuestra propia experiencia nos ofrece el material básico para nuestra imaginación, cuyo alcance es, por lo tanto, limitado. No nos serviría tratar de imaginar que tenemos membranas en nuestros brazos que nos permiten volar entre el crepúsculo y el amanecer cazando insectos con la boca; que tenemos una vista muy pobre, que percibimos el mundo que nos rodea mediante un sistema de señales de sonido reflejadas de alta frecuencia y que nos pasamos el día en un desván colgando boca abajo suspendidos de los pies. En la medida en que puedo imaginar todo esto (que no es mucho), sólo sé cómo sería para *mí* comportarme como se comporta un murciélago. Pero ésa no es la cuestión. Quiero saber cómo es ser murciélago para un *murciélago*. Pero si intento imaginar eso, me veo limitado a los recursos de mi propia mente, y éstos son inadecuados para la tarea. No puedo llevarla a cabo tampoco imaginando adiciones a mi experiencia actual, ni imaginando que ciertos segmentos son sustraídos gradualmente de ella, ni imaginando alguna combinación de adiciones, sustracciones y modificaciones.

[5] Con la frase "nuestro propio caso" no quiero decir solamente "mi propio caso", sino más bien las ideas mentalistas que aplicamos sin problema alguno tanto a nosotros mismos como a otros seres humanos.

En la medida en que sólo puedo parecerme y comportarme como una avispa o un murciélago sin cambiar mi estructura fundamental, mis experiencias no se parecerán en nada a las de esos animales. Por otra parte, es dudoso que podamos darle algún significado a la suposición de que yo debería tener la constitución neurofisiológica interna de un murciélago. Aun cuando pudiera transformarme gradualmente en murciélago, nada en mi constitución actual me permite imaginar cómo serían las experiencias de esa etapa futura en mi metamorfosis. La mejor evidencia nos la darían las experiencias de los murciélagos, si tan sólo supiéramos cómo son.

Así que, si la extrapolación a partir de nuestro propio caso está implicada en la idea de cómo es ser un murciélago, la extrapolación tiene que ser imposible de completar. Sólo podemos desarrollar un concepto esquemático de *cómo* es ser eso. Por ejemplo, podemos atribuir *tipos* generales de experiencia sobre la base de la estructura y el comportamiento del animal. Así, describimos el sonar del murciélago como una forma de percepción delantera tridimensional; creemos que los murciélagos sienten algunos tipos de dolor, temor, hambre y placer, y que poseen otros tipos de percepciones más familiares además del sonar. Pero también creemos que estas experiencias tienen en cada caso un carácter subjetivo específico que no podemos concebir. Y si hay vida consciente en otras partes del universo, es probable que algunas de sus formas no puedan describirse ni siquiera en los términos de experiencia más generales a nuestra disposición.[6] (El problema, sin embargo, no se limita a estos casos exóticos, pues se da entre una persona y otra. Por ejemplo, el carácter subjetivo de la experiencia de una persona ciega y sorda de nacimiento no es accesible para mí, y posiblemente el mío tampoco lo sea para ella. Esto no impide que ambos creamos que la experiencia del otro tiene un carácter subjetivo definido.)

Si alguien se inclina a negar que podemos creer en la existencia de estos hechos cuya naturaleza exacta no podemos con-

[6] Por lo tanto, la forma analógica de la expresión "cómo es ser" provoca confusiones. No significa "a qué se *parece* (en nuestra propia experiencia)", sino más bien "cómo es para el sujeto mismo".

cebir, debería reflexionar que al pensar sobre los murciélagos estamos en la misma posición en que los murciélagos inteligentes o los marcianos[7] estarían si se trataran de formar una idea de cómo sería ser nosotros. La estructura de sus propias mentes podría hacer imposible que tuvieran éxito, pero sabemos que estarían equivocados si concluyeran que no hay nada preciso que sea cómo es ser nosotros: que sólo podrían atribuírsenos ciertos tipos generales de estados mentales (quizá la percepción y el apetito serían conceptos comunes a ellos y nosotros; aunque quizá ni eso). Sabríamos que están equivocados si llegaran a sostener esta conclusión escéptica, porque nosotros sabemos cómo es ser nosotros. Y sabemos que si bien esto incluye una enorme cantidad de variaciones y detalles complejos y que no poseemos un vocabulario para describirlo adecuadamente, su carácter subjetivo es, no obstante, muy específico, y en algunos aspectos es descriptible en términos que sólo pueden comprender las criaturas como nosotros. El hecho de que jamás podremos formular con nuestro lenguaje una descripción detallada de la fenomenología de un marciano o de un murciélago no debería llevarnos a desechar como carente de significado la afirmación de que los murciélagos y los marcianos tienen experiencias totalmente comparables, en cuanto a riqueza de detalles, con las nuestras. Sería muy bueno que alguien desarrollara conceptos y una teoría que nos permitieran pensar sobre esas cosas; pero quizá tal comprensión se nos niegue de manera permanente debido a las limitaciones de nuestra naturaleza. Y negar la realidad o la importancia lógica de lo que nunca podremos describir o comprender es la forma más cruda de disonancia cognoscitiva.

Esto nos lleva al borde de un tema que requiere una discusión mucho más amplia que la que puedo ofrecer aquí, a saber, por un lado, la relación entre los hechos y, por el otro, los esquemas conceptuales o sistemas de representación. Mi realismo acerca del dominio subjetivo en todas sus formas implica una creencia en la existencia de hechos fuera del alcance de los conceptos humanos. Sin duda para un ser humano es posible creer que hay

[7] Cualquier ser extraterrestre inteligente totalmente diferente de nosotros.

hechos para cuya representación o comprensión nunca *tendrá* los conceptos necesarios. Ciertamente sería una tontería dudar de esto dada la finitud de las expectativas de la humanidad. Después de todo, habría habido números transfinitos aunque todo el mundo hubiera muerto de peste negra antes de que Cantor los descubriera. Pero también podemos creer que hay hechos que los seres humanos no *podrían* representar o comprender, aun si la especie durara para siempre, simplemente porque nuestra estructura no nos permite manejar conceptos del tipo necesario. Podría suceder incluso que otros seres advirtieran esta imposibilidad, pero no resulta obvio que la existencia de esos seres, o la posibilidad de su existencia, sea una condición previa de la importancia de la hipótesis de que hay hechos humanamente inaccesibles. (Después de todo, la naturaleza de los seres que tienen acceso a los hechos humanamente inaccesibles es probablemente un hecho humanamente inaccesible.) Por lo tanto, la reflexión sobre cómo es ser murciélago parece llevarnos a concluir que hay hechos que no consisten en la verdad de proposiciones expresables en un lenguaje humano. Podemos vernos obligados a reconocer la existencia de tales hechos sin ser capaces de enunciarlos o comprenderlos.

De cualquier modo, no continuaré analizando este tema. Su relación con el tema que estamos estudiando (esto es, el problema mente-cuerpo) es que nos permite hacer una observación general sobre el carácter subjetivo de la experiencia. Cualquiera que sea el estatus de los hechos acerca de cómo es ser humano, murciélago o marciano, parecen ser hechos que plasman un punto de vista particular.

No me refiero aquí a la supuesta privacía de la experiencia del que la posee. El punto de vista en cuestión no es algo accesible sólo a un único individuo. Se trata más bien de un *tipo* de punto de vista. A menudo podemos adoptar un punto de vista que no es el nuestro, por lo que la comprensión de esos hechos no se limita al caso propio. Hay un sentido en el que los hechos fenomenológicos son perfectamente objetivos: una persona puede saber o decir cuál es la cualidad de la experiencia de otra. Sin embargo, estos hechos son subjetivos en el sentido de que incluso esa atribución objetiva de la experiencia sólo es posible para alguien que sea suficientemente similar al objeto

de la atribución como para poder adoptar su punto de vista; para comprender la atribución en la primera persona del singular al igual que en la tercera del singular, por decirlo así. Entre más diferente de uno sea el otro sujeto de experiencias, menos éxito puede esperarse en esa tarea. En nuestro propio caso empleamos el punto de vista pertinente, pero tendríamos la misma dificultad para comprender nuestra propia experiencia adecuadamente si la enfocáramos desde otro punto de vista que si tratáramos de comprender la experiencia de otra especie sin adoptar *su* punto de vista.[8]

Esto se relaciona en forma directa con el problema mente-cuerpo, pues si los hechos de la experiencia –los hechos acerca de cómo es *para* el organismo que los experimenta– sólo son accesibles desde un punto de vista, entonces es un misterio cómo podría revelarse el verdadero carácter de las experiencias en el funcionamiento físico de ese organismo. Esto último es un dominio de hechos objetivos *par excellence*, la clase de hechos que pueden ser observados y comprendidos desde muchos puntos de vista y por individuos que tienen diferentes sistemas perceptuales. No hay obstáculos imaginativos comparables para que los científicos humanos adquieran conocimientos sobre la neurofisiología de un murciélago; y los murciélagos o los marcianos inteligentes podrían aprender más de lo que nosotros podremos aprender nunca acerca del cerebro humano.

[8] Podría ser más fácil de lo que supongo trascender las barreras entre especies con la ayuda de la imaginación. Por ejemplo, los ciegos son capaces de detectar objetos cercanos a ellos con una especie de sonar, utilizando clics vocales o los golpecitos de un bastón. Quizá si uno supiese cómo es eso, podría imaginar toscamente por extensión cómo sería poseer el sonar mucho más refinado de un murciélago. La distancia entre uno mismo y otras personas y otras especies puede encontrarse en cualquier parte a lo largo de un *continuo*. Incluso para las otras personas la comprensión de cómo es ser ellos es sólo parcial, y cuando consideramos especies muy diferentes de uno mismo, podemos disponer de una comprensión parcial en menor grado aún. La imaginación es notablemente flexible. De cualquier modo, mi idea no es que no podamos *saber* cómo es ser un murciélago. No estoy planteando el problema epistemológico; más bien quiero sostener que incluso para formarnos una *concepción* de cómo es ser murciélago (y *a fortiori* saber cómo es ser murciélago) debemos adoptar el punto de vista del murciélago. Si lo podemos adoptar tosca o parcialmente, entonces nuestra concepción sería también tosca o parcial. O al menos así parece en nuestro estado actual de entendimiento.

Esto no es por sí mismo un argumento contra la reducción. Un científico marciano que no comprendiera la percepción visual podría entender el fenómeno del arco iris, de los rayos o de las nubes como fenómenos físicos, aunque nunca pudiera comprender los conceptos humanos de arco iris, rayo o nube, o el lugar que ocupan esas cosas en nuestro mundo fenoménico. Podría comprender la naturaleza objetiva de las cosas designadas por esos conceptos porque, aunque los conceptos mismos están conectados con un punto de vista particular y con una fenomenología visual particular, las cosas aprehendidas desde ese punto de vista no lo están: son observables desde el punto de vista pero son externas a éste. Por lo tanto, pueden ser comprendidas desde otros puntos de vista, ya sea por los mismos organismos o por otros. El rayo tiene un carácter objetivo que no se agota en su apariencia visual, y este aspecto puede ser investigado por un marciano ciego. En términos más precisos, el rayo tiene un carácter *más* objetivo que lo que se revela en su apariencia visual. Al hablar del cambio de una caracterización subjetiva a una objetiva, no deseo comprometerme con la existencia de un punto final, a saber, la naturaleza intrínseca completamente objetiva de la cosa a la que podríamos o no acceder. Podría ser más preciso pensar en la objetividad como en una dirección en la que puede moverse el entendimiento. Y al comprender un fenómeno como el rayo, es legítimo avanzar y alejarnos lo más que podamos de un punto de vista estrictamente humano.[9]

Por otro lado, en el caso de la experiencia la conexión con un punto de vista particular parece mucho más próxima. Es difícil comprender qué puede significar el carácter *objetivo* de una experiencia además de un punto de vista particular desde el cual un sujeto la aprehende. Después de todo, ¿qué quedaría de cómo es ser murciélago si eliminamos el punto de vista del murciélago? Pero si la experiencia no tiene, además de su carácter

[9] Por lo tanto, el problema que voy a plantear puede formularse incluso si la distinción entre descripciones o puntos de vista más subjetivos y más objetivos sólo puede hacerse desde un punto de vista humano más amplio. No acepto este tipo de relativismo conceptual, pero no necesita ser refutado para que podamos sostener que la reducción psicofísica no puede encajar en el modelo de-lo-subjetivo-a-lo-objetivo que es común en otros casos.

subjetivo, una naturaleza objetiva que pueda ser aprehendida desde muchos puntos de vista diferentes, entonces, ¿cómo podemos suponer que un marciano que investiga mi cerebro podría observar procesos físicos que son mis procesos mentales, sólo que desde un punto de vista diferente (como observaría procesos físicos que fueran relámpagos)? En tal caso, ¿cómo podría un fisiólogo humano observarlos desde otro punto de vista?[10]

Parece que nos enfrentamos a una dificultad general acerca de la reducción psicofísica. En otros campos, el proceso de reducción es un paso hacia una mayor objetividad, hacia un punto de vista más exacto de la naturaleza real de las cosas. Ello se logra reduciendo nuestra dependencia de puntos de vista individuales o específicos de la especie hacia el objeto de la investigación. No lo describimos en términos de las impresiones que crea en nuestros sentidos, sino en términos de sus efectos más generales y de las propiedades que pueden detectarse por medios distintos de los sentidos humanos. Cuanto menos dependa de un punto de vista específicamente humano, más objetiva será nuestra descripción. Es posible seguir este camino porque, aunque los conceptos y las ideas que empleamos al reflexionar sobre el mundo externo se aplican inicialmente desde un punto de vista que incluye nuestro aparato perceptual, los empleamos para referirnos a cosas que están más allá de ellos —de las cuales *tenemos* el punto de vista fenoménico. Por lo tanto, podemos abandonar este punto de vista en favor de otro, y seguir reflexionando sobre las mismas cosas.

Sin embargo, la experiencia misma no parece encajar en este patrón. La idea de ir de la apariencia a la realidad no parece tener sentido aquí. ¿Qué analogía podemos usar, en este caso, para buscar una comprensión más objetiva de los mismos fenómenos abandonando el punto de vista subjetivo inicial hacia ellos, en favor de otro más objetivo pero que concierne a la misma cosa? Ciertamente *parece* poco probable que nos acerque-

[10] El problema no es sólo que cuando veo la "Mona Lisa" mi experiencia visual posee cierto carácter, y nadie que viera en mi cerebro podría encontrar algún indicio de él. Pues incluso si el observador viera en mi cerebro una pequeña imagen de la "Mona Lisa", no tendría ninguna razón para identificarla con la experiencia.

mos a la naturaleza real de la experiencia humana si dejamos atrás la particularidad de nuestro propio punto de vista humano y nos esforzamos por lograr una descripción formulada en términos accesibles para seres que no pueden imaginar cómo sería ser nosotros. Si el carácter subjetivo de la experiencia sólo se comprende plenamente desde un punto de vista, entonces cualquier cambio en busca de una mayor objetividad —esto es, menos vinculado con un punto de vista específico— no nos acerca a la naturaleza real del fenómeno; nos aleja de ella.

En cierto sentido, las simientes de esta objeción a la reductibilidad de la experiencia pueden detectarse ya en los casos exitosos de reducción. Al descubrir que el sonido era, en realidad, un fenómeno ondulatorio en el aire o en otros medios, abandonamos un punto de vista para adoptar otro, y el punto de vista auditivo, humano o animal, que dejamos atrás, permanece sin reducir. Los individuos de dos especies radicalmente distintas pueden comprender los mismos sucesos físicos en términos objetivos, y esto no requiere que comprendan las formas fenomenológicas en que esos sucesos se presentan a los sentidos de los miembros de la otra especie. Por lo tanto, es una condición para su referencia a una realidad común que sus puntos de vista más particulares no sean parte de la realidad común que aprehenden los miembros de ambas especies. La reducción sólo puede tener éxito si el punto de vista específico de cada especie se omite en lo que va a ser reducido.

Sin embargo, aunque tengamos razón en hacer a un lado este punto de vista al buscar una comprensión más plena del mundo externo, no podemos ignorarlo permanentemente, pues es la esencia del mundo interno y no simplemente un punto de vista sobre éste. En gran parte, el neoconductismo de la reciente psicología filosófica es resultado del esfuerzo por sustituir la cosa real por un concepto objetivo de la mente, de manera que no quede nada que no pueda reducirse. Si reconocemos que una teoría física de la mente debe explicar el carácter subjetivo de la experiencia, debemos admitir que ninguna concepción actual nos ofrece una pista sobre cómo podría hacerse eso. Este problema es único. Si los procesos mentales son realmente pro-

cesos físicos, entonces hay algo que es, intrínsecamente,[11] cómo es experimentar ciertos procesos físicos. Cómo podría ser esto continúa siendo un misterio.

¿Qué lección debemos sacar de estas reflexiones y qué debemos hacer después? Sería un error concluir que el fisicalismo debe ser falso. Lo inadecuado de las hipótesis fisicalistas que

[11] Por lo tanto, la relación no sería contingente como en el caso de una causa y su efecto. Sería necesariamente verdadero que cierto estado físico se sintiera de cierto modo. Kripke (1972) sostiene que el conductismo causal y otros análisis relacionados de lo mental fracasan porque interpretan, por ejemplo, "dolor", como un nombre simplemente contingente de los dolores. El carácter subjetivo de la experiencia ("su carácter fenomenológico inmediato", como Kripke lo llama [p. 340]) es la propiedad esencial que tales análisis dejan fuera, y aquella en virtud de la cual es, necesariamente, la experiencia que es. Mi postura es cercana a la suya. Al igual que Kripke, creo que la hipótesis de que cierto estado cerebral debería tener *necesariamente* cierto carácter subjetivo es incomprensible si no se ofrece una mayor explicación. Esta explicación no aparece en las teorías que consideran *contingente* la relación mente-cerebro, pero quizá existan otras alternativas aún por descubrirse.

Una teoría que explicara cómo la relación mente-cerebro es necesaria nos dejaría aún con el problema de Kripke de explicar por qué de todos modos parece contingente. Considero que tal dificultad puede superarse de la siguiente forma. Podemos imaginar algo representándonoslo ya sea de manera perceptual, empática o simbólica. No intentaré explicar cómo funciona la imaginación simbólica, pero una parte de lo que sucede en los otros dos casos es lo siguiente. Para imaginar algo perceptualmente nos situamos en un estado consciente que se parezca al estado en el que nos encontraríamos si lo percibiéramos. Al imaginar algo empáticamente, nos colocamos en un estado consciente que se parezca a la cosa misma. (Este método sólo puede ser usado para imaginar sucesos y estados mentales, los nuestros o los de alguien más.) Cuando tratamos de imaginar un estado mental como si tuviera lugar sin su estado cerebral asociado, imaginamos en primer lugar simpatéticamente que el estado mental ocurre: es decir, nos colocamos en un estado que se parece a él mentalmente. Al mismo tiempo, tratamos de imaginar perceptualmente que no ocurre el estado físico asociado colocándonos en otro estado desconectado del primero: uno que se parece al estado en el que nos encontraríamos si percibiéramos que no ocurre el estado físico. Cuando la imaginación de las características físicas es perceptual y la imaginación de las características mentales es simpatética, nos parece que podemos imaginar cualquier experiencia como si tuviera lugar sin su estado cerebral asociado, y viceversa. La relación entre ellos parecerá contingente, incluso si fuese necesaria, debido a la independencia de los diferentes tipos de imaginación.

(El solipsismo, dicho sea de paso, aparece si malinterpretamos la imaginación simpatética como si funcionara como la imaginación perceptual: parece entonces imposible imaginar cualquier experiencia que no sea propia.)

presuponen un análisis objetivo defectuoso de la mente no prueba nada. Sería más verdadero afirmar que el fisicalismo es una tesis que no podemos comprender porque no tenemos a la fecha una idea de cómo podría ser verdadero. Quizá se considere poco razonable exigir que tal concepción sea una condición de la comprensión. Después de todo, podría decirse, el significado del fisicalismo está muy claro: los estados mentales son estados del cuerpo; los sucesos mentales son sucesos físicos. No sabemos *cuáles* son estos estados y sucesos físicos, pero esto no debería impedirnos comprender la hipótesis. ¿Qué puede estar más claro que las palabras "es" y "son"?

Sin embargo, creo que es precisamente la aparente claridad de la palabra "es" lo que resulta engañoso. Por lo general, cuando nos dicen que *X* es *Y*, sabemos *cómo* se supone que esto es verdadero, pero ello depende de antecedentes conceptuales o teóricos y no se comprende sólo con la palabra "es". Sabemos cómo refieren "*X*" y "*Y*" y el tipo de cosas a las que refieren, y tenemos una idea tosca de cómo las dos rutas referenciales podrían converger en una sola cosa, ya sea un objeto, una persona, un proceso, un suceso o cualquier otra cosa. Pero cuando los dos términos de la identificación son muy distintos, podría no estar muy claro cómo podría ser verdadera. Podríamos no tener siquiera una idea tosca de cómo podrían converger las dos rutas referenciales, o en qué tipo de cosas podrían converger, y quizá necesitemos un marco teórico que nos permita comprender eso. Sin tal marco, la identificación queda rodeada de un aire de misticismo.

Esto explica el sabor mágico de las divulgaciones populares de los descubrimientos científicos fundamentales, que se nos ofrecen como propuestas que debemos aceptar aun sin comprenderlas realmente. Por ejemplo, hoy día a las personas se les dice a temprana edad que toda materia es en realidad energía. Pero a pesar de que saben lo que significa la palabra "es", la mayoría nunca se forma una idea de qué hace verdadero ese enunciado, porque carecen de los antecedentes teóricos.

Hoy día, la postura del fisicalismo es similar a la que habría tenido la hipótesis de que la materia es energía si la hubiera expresado un filósofo presocrático. No tenemos ni el esbozo de

una concepción sobre cómo podría ser verdadera. Para comprender la hipótesis de que un suceso mental es un suceso físico necesitamos algo más que entender la palabra "es". Nos falta la idea de cómo un término físico y uno mental pueden referirse a la misma cosa, y las analogías comunes con identificaciones teóricas en otros campos no han logrado proporcionárnosla. Y no lo han logrado porque, si interpretamos la referencia de los términos mentales a sucesos físicos con el modelo usual, entonces reaparecen sucesos subjetivos separados como los efectos por medio de los cuales se asegura la referencia mental a los sucesos físicos, o bien obtenemos una explicación falsa de cómo refieren los términos mentales (por ejemplo, una explicación conductista causal).

Aunque resulte extraño, podemos tener pruebas de la verdad de algo que realmente no podemos comprender. Supóngase que alguien que no conoce la metamorfosis de los insectos encierra una oruga en una caja fuerte esterilizada y que semanas más tarde abre la caja y encuentra una mariposa. Si esa persona sabe que la caja fuerte estuvo cerrada todo el tiempo, tiene razones para creer que la mariposa es, o fue alguna vez, la oruga, sin tener la menor idea de en qué sentido pudo ocurrir eso. (Una posible explicación sería que la oruga hubiera tenido dentro un parásito pequeño y alado que la hubiera devorado y hubiera crecido hasta convertirse en la mariposa.)

Es posible que estemos en una situación similar en relación con el fisicalismo. Donald Davidson ha argumentado que si los sucesos mentales tienen causas y efectos físicos, deben tener descripciones físicas. Sostiene que tenemos razones para creer esto aunque no tengamos, y de hecho no *podríamos* tener, una teoría psicofísica general.[12] Su argumento se aplica a los sucesos mentales intencionales, pero creo que también tenemos ciertas razones para creer que las sensaciones son procesos físicos, sin estar en condiciones de comprender tal afirmación. La postura de Davidson es que ciertos sucesos físicos tienen, irreductiblemente, propiedades mentales, y quizá alguna postura que pueda ser descrita de esta manera sea correcta. Pero nada que podamos

[12] Véase Davidson 1970; aunque no entiendo el argumento en contra de las leyes psicofísicas.

concebir ahora corresponde a ella, ni tampoco tenemos idea alguna de cómo sería una teoría que nos permitiera concebirla.[13]

Se ha trabajado muy poco en la cuestión básica (en la cual se puede evitar totalmente mencionar el cerebro) de si puede entenderse la idea de experiencias que tienen un carácter objetivo. En otras palabras, ¿tiene sentido preguntar cómo son *realmente* mis experiencias a diferencia de cómo me parecen? No podemos comprender realmente la hipótesis de que su naturaleza se capta en una descripción física a menos que comprendamos la idea más fundamental de que *tienen* una naturaleza objetiva (o que los procesos objetivos pueden tener una naturaleza subjetiva).[14]

Me gustaría terminar con una propuesta especulativa. Quizá sea posible enfocar desde otra dirección el abismo entre lo subjetivo y lo objetivo. Dejando a un lado temporalmente la relación entre la mente y el cerebro, podemos buscar una comprensión más objetiva de lo mental por sí mismo. Hoy día estamos muy mal preparados para pensar en el carácter subjetivo de la experiencia sin apoyarnos en la imaginación (sin considerar el punto de vista del sujeto de la experiencia). Esto debiera considerarse como un reto para crear nuevos conceptos y concebir un nuevo método: una fenomenología objetiva que no dependa de la empatía o de la imaginación. Aunque probablemente no lo abarcaría todo, su objetivo sería describir, por lo menos en parte, el carácter subjetivo de las experiencias en una forma comprensible para los seres incapaces de tener esas experiencias.

Tendríamos que desarrollar tal fenomenología para describir las experiencias sonares de los murciélagos; pero también sería posible comenzar con los seres humanos. Por ejemplo, podríamos intentar desarrollar conceptos que pudieran usarse para explicarle a un ciego de nacimiento cómo es ver. Tarde

[13] Se pueden hacer observaciones similares a mi artículo "Physicalism" (Nagel 1965).

[14] Esta cuestión se encuentra también en el corazón del problema de las otras mentes, cuya relación cercana con el problema mente-cuerpo a menudo se pasa por alto. Si entendiéramos cómo la experiencia subjetiva puede tener una naturaleza objetiva, entenderíamos la existencia de sujetos distintos de uno mismo.

o temprano llegaríamos a un callejón sin salida, pero debe ser posible idear un método para expresar en términos objetivos, y con mayor precisión, mucho más de lo que podemos expresar hoy día. Las analogías intermodales vagas, como por ejemplo "el rojo es como el sonido de una trompeta", que surgen cuando se discute este tema, son de poca utilidad. Esto debería ser evidente para cualquiera que haya escuchado una trompeta y haya visto el color rojo. Pero las características estructurales de la percepción podrían prestarse más a una descripción objetiva, aunque algo se deje afuera. Y ciertos conceptos alternativos distintos de los que aprendemos en primera persona podrían permitirnos alcanzar cierto tipo de comprensión incluso de nuestra propia experiencia, que se nos niega por la misma facilidad de la descripción y la falta de perspectiva propia de los conceptos subjetivos.

Además de su interés intrínseco, una fenomenología que sea objetiva en este sentido podría permitir que las preguntas acerca de la base física[15] de la experiencia adoptaran una forma más inteligible. Los aspectos de la experiencia subjetiva que pudieran adoptar este tipo de descripción objetiva podrían ser mejores candidatos a explicaciones objetivas más conocidas. Pero, sea correcta o no esta suposición, parece improbable que una teoría física de la mente pueda vislumbrarse hasta que no se haya trabajado más en el problema general de lo subjetivo y lo objetivo. De otra forma, no podemos siquiera plantear el problema mente-cuerpo sin que se nos escape.

[*Traducción de Hector Islas*]

[15] No he definido el término "físico". Obviamente no se aplica sólo a lo que puede ser descrito por los conceptos de la física contemporánea, puesto que aún esperamos desarrollos futuros. Algunos piensan que nada impide que los fenómenos mentales sean reconocidos con el tiempo como fenómenos físicos por derecho propio. Pero independientemente de qué más pueda decirse de lo físico, debe ser objetivo. De manera que, si algún día nuestra idea de lo físico se expande hasta incluir a los fenómenos mentales, tendrá que asignarles un carácter objetivo, al margen de si esto se realiza analizándolos en términos de otros fenómenos considerados como físicos. Me parece más probable, sin embargo, que las relaciones entre lo mental y lo físico se formulen finalmente en una teoría cuyos términos fundamentales no pueden clasificarse claramente en ninguna de esas categorías.

BIBLIOGRAFÍA

Armstrong, D.M., 1968, *A Materialist Theory of the Mind*, Routledge and Kegan Paul, Londres.

Capitan, W.H. y D.D. Merrill (comps.), 1967, *Art, Mind, and Religion*, University of Pittsburgh Press, Pittsburgh.

Davidson, D., 1970, "Mental Events", en Foster y Swanson 1970, pp. 79–101. [Versión en castellano: "Sucesos mentales", en *Ensayos sobre acciones y sucesos*, trad. O. Hansberg, J.A. Robles y M.M. Valdés, Crítica/Instituto de Investigaciones Filosóficas-UNAM, Barcelona, 1995, pp. 263–287.]

——— y G. Harman (comps.), 1972, *Semantics of Natural Language*, Reidel, Dordrecht.

Dennett, D.C., 1969, *Content and Consciousness*, Routledge and Kegan Paul, Londres. [Versión en castellano: *Contenido y conciencia*, trad. J.M. Lebrón, Gedisa, Barcelona, 1996.]

Foster, L. y J.W. Swanson (comps.), 1970, *Experience and Theory*, University of Massachussetts Press, Amherst.

Kripke, S., 1972, "Naming and Necessity", en Davidson y Harman 1972, pp. 253–355. (Versión en libro: *Naming and Necessity*, Harvard University Press, Cambridge, Mass., 1980. [Versión en castellano: *El nombrar y la necesidad*, 2a. ed., trad. M.M. Valdés, Instituto de Investigaciones Filosóficas-UNAM, 1995.]

Lewis, D.K., 1966, "An Argument for the Identity Theory", *Journal of Philosophy*, vol. 63, no. 1, pp. 17–25 (reimpreso con una adenda en Rosenthal 1971, pp. 162–171).

Nagel, T., 1971, "Brain Bisection and the Unity of Consciousness", *Synthese*, vol. 22, pp. 396–413.

———, 1970, "Armstrong on the Mind", *Philosophical Review*, vol. 79, no. 3, pp. 394–403.

———, 1965, "Physicalism", *Philosophical Review*, vol. 74, no. 3, pp. 339–356. (Reimpreso con un epílogo en O'Connor 1969, pp. 99-116).

O'Connor, J. (comp.), 1969, *Modern Materialism: Readings on Mind-Body Identity*, Harcourt, Brace and World, Nueva York.

Putnam, H., 1967, "Psychological Predicates", en Capitan y Merrill 1967, pp. 37–48. (Reimpreso en Rosenthal 1971 como "The Nature of Mental States", pp. 150–161.)

Rorty, R., 1965, "Mind-Body Identity, Privacy, and Categories", *The Review of Metaphysics*, vol. 19, no. 1, pp. 24–54.

Rosenthal, D.M. (comp.), 1971, *Materialism and the Mind-Body Problem*, Prentice-Hall, Englewood Cliffs.

Smart, J.J.C., 1963, *Philosophy and Scientific Realism*, Routledege and Kegan Paul, Londres.
Thornton, M.T., 1972, "Ostensive Terms and Materialism", *The Monist*, vol. 56, no. 2, pp. 193–214.

¿PODEMOS RESOLVER EL PROBLEMA MENTE-CUERPO?*

COLIN MCGINN

> El que algo tan notable como un estado de conciencia se produzca como resultado de la irritación del tejido nervioso es tan inexplicable como la aparición del genio cuando Aladino frotó su lámpara en el cuento...
>
> Julian Huxley

Desde hace mucho tiempo hemos tratado de resolver el problema mente-cuerpo. Una y otra vez nuestros mejores esfuerzos se han visto frustrados. El misterio persiste. Creo que ha llegado el momento de admitir con franqueza que no podemos resolverlo. Pero pienso también que esta misma falta de solución —o la razón que la explica— elimina el problema filosófico. En este artículo explico por qué digo estas cosas tan extravagantes.

El problema específico que deseo discutir se refiere a la conciencia, el meollo del problema mente-cuerpo. ¿Cómo es posible que los estados conscientes dependan de los estados del cerebro? ¿Cómo puede surgir una fenomenología multicolor de la opaca materia gris? ¿Qué es lo que hace que el órgano que llamamos cerebro sea tan radicalmente distinto de los otros

* Varias discusiones con las personas que se mencionan a continuación me han ayudado a desarrollar las ideas de este artículo: Anita Avramides, Jerry Katz, Ernie LePore, Michael Levin, Thomas Nagel, Galen Strawson, Peter Unger. Mi gran deuda con la obra de Nagel es algo que debería resultar obvio a lo largo del artículo: no habría intentado enfrentarme al problema mente-cuerpo si él no lo hubiera hecho antes.

órganos del cuerpo como, por ejemplo, los riñones –las partes del cuerpo sin ningún rastro de conciencia? ¿Cómo puede el agregado de millones de neuronas individualmente insensibles generar la experiencia subjetiva? Sabemos que el cerebro es la base causal *de facto* de la conciencia, pero, al parecer, no tenemos idea sobre cómo puede esto ser así. Consideramos que es algo milagroso, misterioso, incluso vagamente cómico. Creemos que, de algún modo, el agua del cerebro físico se convierte en el vino de la conciencia, pero no tenemos ni idea de la naturaleza de esta conversión. Los disparos neuronales aparentemente son el material equivocado para traer la conciencia al mundo; pero, de algún modo, pareciera que son ellos quienes realizan este acto misterioso. El problema mente-cuerpo es el problema de entender cómo se realiza el milagro, eliminando así su sentido de profundo misterio. Queremos quitarle la magia al vínculo entre la conciencia y el cerebro.[1]

Las supuestas soluciones al problema han tendido a adoptar una de dos formas. Una de ellas, que podemos llamar constructiva, intenta especificar alguna propiedad natural del cerebro (o del cuerpo) que explique cómo puede producirse la conciencia a partir de ella. Así, el funcionalismo, por ejemplo, sugiere que existe una propiedad –a saber, el papel causal– que satisfacen tanto los estados cerebrales como los estados mentales; esta propiedad supuestamente explica cómo los estados conscientes

[1] Una de las peculiaridades del problema mente-cuerpo es la dificultad para formularlo en una forma rigurosa. Nuestra visión del problema rebasa nuestra capacidad de articularlo con claridad. Así, nos encontramos de pronto recurriendo a invitaciones a mirar hacia adentro, en vez de especificar con precisión *qué* es lo que convierte la conciencia en algo inexplicable en términos de propiedades físicas ordinarias. Y esto puede producir la impresión de que el problema es espurio. Una criatura sin conciencia no tendría una idea adecuada del problema (suponiendo que tal criatura pudiera reconocer otros problemas). Pienso que un tratamiento adecuado del problema mente-cuerpo debería explicar por qué es tan difícil formular el problema de una manera explícita. Mi tratamiento ubica la dificultad en nuestras concepciones inadecuadas de la naturaleza del cerebro y de la conciencia. De hecho, si conociéramos cabalmente sus naturalezas, ya habríamos resuelto el problema. Espero que esto se aclare más adelante.

provienen de los estados del cerebro.[2] La otra forma que ha dominado a lo largo de la historia, admite con franqueza que nada que sea simplemente natural puede realizar el trabajo, y sugiere entonces que invoquemos entidades sobrenaturales o intervenciones divinas. Así es como obtenemos un dualismo cartesiano o una armonía preestablecida leibniziana. Estas "soluciones" al menos reconocen que se necesita algo fuera de lo común para darle sentido a la relación mente-cuerpo; ellas son tan extremas como el problema mismo. El enfoque que yo favorezco es naturalista pero a la vez no constructivo: creo que nunca vamos a poder especificar qué cosa del cerebro es responsable de la conciencia, pero estoy seguro de que, sea lo que sea, no es algo inherentemente milagroso. Sugiero que el problema surge porque nuestra propia constitución cognoscitiva hace imposible una concepción de esa propiedad natural del cerebro (o de la conciencia) que explica el eslabón psicofísico. Se trata de un nexo causal que nunca podremos comprender dada la manera que tenemos de formar nuestros conceptos y de desarrollar nuestras teorías. ¡Con razón el problema nos parece tan difícil!

Para poder formular de una manera plausible esta idea, necesito bosquejar la concepción general de competencia cognoscitiva que sustenta mi postura. Permítaseme introducir la idea de *clausura cognoscitiva*. Un tipo de mente M está cognoscitivamente cerrado con respecto a la propiedad P (o la teoría T) si y sólo si los procedimientos de formación de conceptos a disposición de M no alcanzan a comprender P —o a comprender T. Hay diferentes tipos de mentes con capacidad cognoscitiva, equipadas con distintos poderes y limitaciones, propensiones y puntos ciegos, de manera tal que las propiedades (o teorías)

[2] También incluiría el panpsiquismo dentro de la categoría de las soluciones constructivas, ya que intenta explicar la conciencia en términos de propiedades del cerebro que son tan naturales como la conciencia misma. Atribuir partículas de protoconciencia a los constituyentes de la materia no es sobrenatural del modo como sí lo es postular sustancias inmateriales o intervenciones divinas; es simplemente extravagante. Supondré aquí que el panpsiquismo, al igual que todas las otras soluciones constructivas existentes, es inadecuado como respuesta al problema mente-cuerpo —así como también lo son, desde luego, las "soluciones" sobrenaturales. Me dirijo aquí a aquellos que aún se sienten perplejos (casi todos diría yo, al menos en el fondo).

pueden ser accesibles para algunas mentes pero no para otras. Lo que está cerrado para la mente de una rata puede estar abierto para la mente de un mono, y lo que está abierto para nosotros puede estar cerrado para el mono. La capacidad representacional no es un asunto de todo o nada. Las mentes son productos biológicos tal como los cuerpos y, al igual que ellos, las hay de diferentes tamaños y formas, más o menos capaces, más o menos adecuadas para realizar determinadas tareas cognoscitivas.[3] Esto, desde luego, es particularmente claro en lo que respecta a las facultades de percepción: la clausura perceptual es algo que difícilmente puede negarse. Hay diferentes especies que son capaces de percibir diferentes propiedades del mundo, y ninguna especie es capaz de percibir todas las propiedades que las cosas pueden ejemplificar (por lo menos no sin utilizar instrumentos artificiales). Pero tal clausura no disminuye la realidad de las propiedades que se encuentran más allá de las capacidades representacionales en cuestión; una propiedad no es menos real por el hecho de encontrarse fuera del alcance de cierto tipo de mente percipiente con capacidad cognoscitiva. Las partes invisibles del espectro electromagnético son tan reales como sus partes visibles, y el hecho de que un tipo específico de criatura pueda o no formar representaciones conceptuales de esas partes imperceptibles no determina el hecho de que existan o no. Por lo tanto, la clausura cognoscitiva con respecto a P no implica un irrealismo sobre P. Que P sea (como puede decirse) *nouménico* para M no muestra que P no aparezca en alguna teoría científica naturalista T; sólo muestra que T no es accesible cognoscitivamente para M. Quizá las mentes de los monos y la propiedad de ser un electrón ejemplifican esta posibilidad. Y aquí necesariamente debe surgir la pregunta de si las mentes humanas se encuentran cerradas con respecto a de-

[3] Este tipo de concepción de capacidad cognoscitiva es defendido enérgicamente por Noam Chomsky 1975, y por Jerry Fodor 1983. Chomsky traza una distinción entre "problemas", para cuya solución las mentes humanas están en principio equipadas, y "misterios", los cuales eluden sistemáticamente nuestro entendimiento. El autor concibe además un estudio de nuestros sistemas cognoscitivos que clasificaría esos poderes y limitaciones. Yo estoy comprometido aquí con tal estudio, y considero que el problema mente-cuerpo debe clasificarse entre los misterios.

terminadas teorías explicativas verdaderas. Al menos nada en el concepto de realidad muestra que todo lo real esté abierto para la facultad humana de formar conceptos; esto es, si somos realistas acerca de la realidad.[4]

Consideremos una mente construida de acuerdo con los principios del empirismo clásico; una mente humeana. Hume pensó, equivocadamente, que las mentes humanas eran humeanas, pero podemos al menos concebir tal tipo de mente (quizá los perros y los monos poseen mentes humeanas). En una mente humeana la clausura perceptual determina la clausura cognoscitiva, ya que las "ideas" deben ser siempre copias de "impresiones"; por lo tanto, el sistema de formación de conceptos no puede trascender lo que puede ser presentado al sujeto en términos perceptuales. Una mente con esas características estará cerrada con respecto a los entes inobservables; por ejemplo, las propiedades de los átomos no serán representables por una mente construida de esa manera. Esto implica que las teorías explicativas donde figuren de una manera básica estas propiedades no serán accesibles para una mente humeana.[5] Por lo tanto, los fenómenos observables que se explican mediante la referencia a entes inobservables serán inexplicables para una mente con tales limitaciones. Pero nótese bien: la incapacidad de explicar

[4] Véase la discusión de Thomas Nagel sobre el realismo en Nagel 1986, cap. VI. El autor argumenta ahí en favor de la posibilidad de la existencia de propiedades que nunca podremos comprender. Si combinamos el realismo de Nagel con la clausura cognoscitiva de Chomsky-Fodor obtenemos una posición que se parece mucho a la de Locke en *Ensayo sobre el entendimiento humano*: la idea de que nuestras facultades dadas por Dios no nos capacitan para comprender la verdad profunda de la realidad. De hecho, Locke sostuvo precisamente esto en lo que respecta a la relación entre la mente y el cerebro: sólo la revelación divina nos permitiría entender cómo se producen las "percepciones" en nuestras mentes mediante los objetos materiales.

[5] Desde luego, Hume argumentaba en efecto que ninguna teoría que empleara de una manera básica una noción de necesidad objetiva causal podría ser comprendida por nuestras mentes –y lo mismo vale para la noción de persistencia objetiva. Podemos comparar las frustraciones de la mente humeana con las congojas conceptuales de los seres de sonido puro que se discuten en el segundo capítulo de *Individuals* de Strawson (1959); ambas son tipos de mente cuya constitución coloca varios conceptos fuera de su alcance. Nosotros podemos desempeñarnos mucho mejor que estas mentes truncas, pero también tenemos nuestras limitaciones conceptuales.

ciertos fenómenos no conlleva una falta de reconocimiento de los problemas teóricos que los fenómenos plantean. Podemos ser capaces de apreciar un problema sin ser capaces de formular (ni siquiera en principio) la solución de ese problema (supongo que los niños se encuentran a menudo en una posición semejante, al menos por un rato). Una mente humeana no puede resolver los problemas que nuestros físicos resuelven, pero podría tener una idea de lo que requiere explicación. Podríamos esperar, entonces, que una mente inquisitiva humeana moderadamente inteligente se sintiera permanentemente perpleja y desconcertada ante el mundo físico, ya que la ciencia correcta se encuentra para siempre más allá de su alcance cognoscitivo. De hecho, la posición de Locke era precisamente algo parecido a ésta. Locke pensaba que nuestras ideas de la materia estaban restringidas de una manera muy precisa por nuestras percepciones, y por tanto, concluía que la verdadera ciencia de la materia se encuentra eternamente más allá de nuestro alcance; que nunca podríamos eliminar nuestras perplejidades frente a, por ejemplo, lo que esencialmente es la solidez.[6] Sin embargo, para Locke no se sigue de lo anterior que la naturaleza sea en sí misma inherentemente misteriosa; el carácter de misterio que experimentamos proviene de nuestras propias limitaciones cognoscitivas, y no de algún misterio objetivo en el mundo. Hoy nos parece que Locke estaba equivocado acerca de nuestra capacidad para comprender la naturaleza del mundo físico, pero aún podemos aprender algo de esta idea fundamental —la insistencia en que nuestras facultades cognoscitivas podrían no ser capaces de resolver todos los problemas con que nos topemos. Dicho en términos más generales: puede ser que la mente humana no se ajuste a los principios empiristas, pero debe ajustarse a *algunos* principios —y es una afirmación muy importante decir que estos principios permiten la solución de todo problema que podamos formular o intuir. Una apertura cognoscitiva total no es algo que esté garantizado para los seres humanos, y no debería esperarse. Pero lo que es nouménico para nosotros podría

[6] Véase su *Ensayo sobre el entendimiento humano*, Libro II, cap. IV. Locke compara la tarea de explicar lo que es esencialmente la solidez con el intento de aclarar la visión de un hombre ciego hablando con él.

no ser milagroso en sí mismo. Deberíamos, por lo tanto, estar alerta ante la posibilidad de que un problema que nos parezca profundamente inextricable, absolutamente desconcertante, pueda surgir de un área de clausura cognoscitiva en nuestras formas de representar el mundo.[7] Lo que quiero argumentar ahora es que esto es lo que sucede con nuestra intuición de la naturaleza misteriosa de la conexión de la conciencia con el cerebro. Tendemos a desviarnos de la teoría explicativa correcta del nexo psicofísico, y ello nos hace propensos a aceptar la ilusión de un misterio objetivo. La apreciación de esto debería eliminar el problema filosófico: en realidad, la conciencia no surge del cerebro de la manera misteriosa en que el genio surge de la lámpara.

Ahora necesito argumentar en favor de tres cosas: (i) existe una propiedad del cerebro que explica la conciencia en términos naturalistas; (ii) nosotros estamos cerrados cognoscitivamente a esa propiedad; pero (iii) no existe un problema mente-cuerpo filosófico (en oposición a un problema mente-cuerpo científico). La mayor parte de mi trabajo está consagrado a argumentar en favor de (iii).

Una vez que hemos rehuido resueltamente lo sobrenatural, pienso que es innegable que debe ser en virtud de *alguna* propiedad natural del cerebro que los organismos son conscientes. Simplemente *tiene* que haber una explicación de cómo los cerebros promueven la conciencia. Si no hemos de ser eliminativistas con respecto a la conciencia, entonces tenemos que pensar que existe una teoría que explique las correlaciones psicofísicas que observamos. No es plausible adoptar una posición en la que estas correlaciones se consideren como hechos últimos e inexplicables; como simples hechos brutos. Tampoco querríamos aceptar la emergencia radical de la conciencia con respecto al cerebro: eso se parece demasiado a aceptar milagros *de re*. Sabe-

[7] Podría pensarse que algunos de los aspectos más misteriosos de la cosmología y la teoría cuántica se encuentran justo dentro de los límites de la inteligibilidad humana. Chomsky sugiere que el carácter causal de la conducta podría ser algo necesariamente misterioso para los investigadores humanos: véase Chomsky 1975, p. 156. Yo mismo creo que el problema mente-cuerpo manifiesta un nivel de misterio cualitativamente distinto de ese caso (a menos que se tome como un aspecto de ese problema).

mos que los estados cerebrales causan los estados conscientes, y este nexo causal debe darse mediante conexiones necesarias de algún tipo —del tipo que haría inteligible al nexo *si* se entendieran.[8] En este aspecto la conciencia es como la vida. Sabemos que la vida se desarrolló a partir de la materia inorgánica, por lo que esperamos que haya alguna explicación de este proceso. No es posible que consideremos el nacimiento de la vida como un hecho primitivo y bruto, ni tampoco podemos aceptar que la vida haya surgido por alguna forma milagrosa. Más bien, debe haber una explicación natural de cómo puede surgir la vida a partir de la materia, independientemente de que podamos conocerla o no. Evadiendo el vitalismo y el toque mágico del dedo de Dios, insistimos correctamente en que debe ser en virtud de alguna propiedad natural de la materia (organizada) que ciertas porciones de ella se convierten en algo vivo. Pero la conciencia en sí no es otra cosa que un desarrollo biológico más, y por tanto también debe ser susceptible de alguna explicación natural —independientemente de que los seres humanos sean capaces de llegar a ella. Suponemos que existen leyes naturales objetivas que de algún modo explican el surgimiento repentino de la conciencia. En pocas palabras, la conciencia tiene que ser un fenómeno natural, que surge de manera natural a partir de ciertas formas de organización de la materia. Digamos entonces que existe una propiedad P, instanciada en el cerebro, en virtud de la cual este órgano es la base de la conciencia. De manera equivalente, existe una teoría T, que hace referencia a P, que explica de manera completa la dependencia de los estados conscientes con respecto a los estados cerebrales. Si conociéramos T, entonces tendríamos una solución constructiva para el problema mente-cuerpo. La pregunta es entonces si podemos o no conocer T y comprender la naturaleza de P.

Permítaseme mencionar primero que sin duda es *posible* que nunca podamos llegar a comprender P; no hay, como he dicho, garantía alguna de que nuestras capacidades cognoscitivas nos permitan resolver todo problema que podamos reconocer

[8] *Cfr.* la discusión de Nagel sobre la emergencia en "Panpsychism", en Nagel 1979. Estoy de acuerdo con él en que la aparente emergencia radical de la mente a partir de la materia tiene que ser meramente epistémica, so pena de aceptar milagros inexplicables en el mundo.

como tal. Sólo un idealismo desmedido sobre el mundo natural podría garantizar la afirmación dogmática de que todo es cognoscible por la especie humana en la etapa presente de su desarrollo evolutivo (considérese la misma afirmación en relación con el intelecto del hombre de Cro-Magnon). *Podría* ser que toda propiedad para la cual podamos formar un concepto sea tal que *ella* no pueda nunca resolver el problema mente-cuerpo. *Podríamos* ser como niños de cinco años tratando de entender la teoría de la relatividad. Con todo, hasta ahora esto sólo es una afirmación de posibilidad: ¿qué razón tenemos para afirmar, de manera positiva, que nuestras mentes se hallan cerradas con respecto a *P*?

La larga serie de fracasos en la historia es sugerente, pero está lejos de ser concluyente. ¿Quizá, se diría, la solución se encuentra a la vuelta de la esquina?, o ¿tiene que esperar que las ciencias físicas estén completas? ¿Quizá simplemente falta que produzcamos un genio como Einstein que reestructure el problema de una manera astuta y presente entonces la solución a un mundo atónito?[9] Sin embargo, pienso que nuestro profundo desconcierto ante el problema, que conduce a un vertiginoso sentido de misterio fundamental, y que se resiste incluso a una formulación clara, debería al menos motivarnos a explorar la idea de que existe algo terminal con respecto a nuestra perplejidad. Así como los teólogos tradicionales tuvieron que aceptar una clausura cognoscitiva en relación con ciertas propiedades de Dios, así nosotros deberíamos considerar en serio la idea de que el problema mente-cuerpo nos lleva a chocar contra los límites de nuestra capacidad para comprender el mundo. Eso es justo lo que haré ahora.

Parece que hay dos caminos posibles para nuestra aspiración de identificar *P*: podríamos intentar alcanzar *P* investigando

[9] A pesar de su reputación de pesimista con respecto al problema mente-cuerpo, una lectura cuidadosa de Nagel revela una vena optimista en su pensamiento (medido con los parámetros del presente artículo); véanse, en particular, las observaciones finales de "What Is It Like to Be a Bat?", en Nagel 1979 ["¿Como es ser un murciélago?" incluido en esta antología, pp. 45-63]. Nagel especula que quizá podamos idear una "fenomenología objetiva" que haga los estados conscientes más manejables para un análisis físico. A diferencia de mi posición, Nagel no considera que el problema esté inherentemente más allá de nuestro alcance.

directamente la conciencia, o podríamos, en nuestra búsqueda de *P*, apelar al estudio del cerebro. Consideremos estas alternativas en orden, empezando por el estudio de la conciencia. Nuestra familiaridad con la conciencia difícilmente podría ser más directa; así, la descripción fenomenológica se da de una manera (relativamente) fácil. La facultad mediante la cual apresamos la conciencia en toda su vívida desnudez se llama "introspección". En virtud de poseer esta facultad cognoscitiva nos podemos atribuir conceptos de conciencia a nosotros mismos; tenemos así un "acceso inmediato" a las propiedades de la conciencia. Pero, ¿acaso la facultad de introspección nos revela la propiedad *P*? ¿Podemos decir sólo mediante la introspección cuál es la solución al problema mente-cuerpo? Evidentemente no. Tenemos acceso cognoscitivo directo a uno de los términos de la relación mente-cuerpo, pero no tenemos el mismo acceso a la naturaleza del vínculo. La introspección no presenta de algún modo inteligible los estados conscientes *como* dependientes del cerebro. No podemos, por lo tanto, conocer *P* mediante la introspección. Más aún, parece imposible que podamos algún día aumentar nuestra provisión de conceptos atribuidos introspectivamente con la incorporación de *P* —es decir, que podamos adquirir este concepto simplemente sobre la base de una introspección cuidadosa y prolongada. La fenomenología pura jamás nos ofrecerá la solución al problema mente-cuerpo. Tampoco parece factible tratar de extraer *P* de los conceptos de la conciencia que actualmente poseemos mediante algún procedimiento de análisis conceptual —ello sería tan poco factible como intentar resolver el problema vida-materia reflexionando simplemente sobre el concepto de *vida*.[10] *P* debe encontrarse fuera del campo de lo introspectible, y tampoco se encuentra

[10] Esta posición es quizá la más notablemente optimista de todas —la esperanza de que una reflexión sobre el concepto ordinario de (por ejemplo) dolor revele la manera en que el dolor depende del cerebro. Si no me equivoco, ésta es en efecto la tesis de los funcionalistas de sentido común: piensan que *P* consiste en un papel causal, y que esto puede inferirse analíticamente a partir de los conceptos de los estados conscientes. Esto haría realmente sorprendente el hecho de que alguna vez hayamos sentido que existe un problema mente-cuerpo, puesto que su solución está ya contenida en nuestros conceptos mentales. ¡Qué optimismo!

contenido implícitamente en los conceptos que aplicamos en nuestras atribuciones en primera persona. Así, la facultad de introspección, en tanto capacidad para formar conceptos, se encuentra cognoscitivamente cerrada con respecto a *P*; lo que no debería sorprendernos en vista de su campo de acción tan limitado (la *mayoría* de las propiedades del mundo están cerradas a la introspección).

Pero existe una idea más por examinar respecto a *P* y a la conciencia, y que tiene que ver con nuestro acceso restringido a los conceptos mismos de conciencia. Es común la afirmación de que la gama de los conceptos de conciencia al alcance de una mente *M* está limitado por las formas específicas de conciencia poseídas por *M*. Dicho de una manera tosca, no podemos formar conceptos de propiedades conscientes a menos que nosotros mismos las instanciemos. Un ciego de nacimiento no puede comprender el concepto de una experiencia visual del rojo, y los seres humanos no pueden concebir las experiencias de ecolocalización de los murciélagos.[11] Éstos son casos de clausura cognoscitiva dentro de la clase de las propiedades conscientes. Pero entonces, al parecer este tipo de clausura afectará nuestras esperanzas de acceder a *P*. Pues supóngase que estuviéramos cognoscitivamente abiertos con respecto a *P*; es decir, supóngase que tuviéramos la solución del problema de cómo ciertas formas específicas de conciencia dependen de diferentes tipos de estructuras fisiológicas. Entonces, entenderíamos por supuesto cómo el cerebro de un murciélago produce las experiencias subjetivas de los murciélagos. Llamemos *B* a este tipo de experiencia, y llamemos P_I a la propiedad explicativa que vincula a *B* con el cerebro del murciélago. Mediante la comprensión de P_I, el fenómeno de cómo el cerebro del murciélago genera las experiencias *B* sería perfectamente inteligible

[11] Véase Nagel 1979. Adviértase que el carácter fugaz de tales propiedades con respecto a nuestros conceptos nada tiene que ver con su "complejidad". Al igual que las propiedades fugaces de los colores, tales propiedades de la experiencia son "simples". Adviértase además que tales propiedades ofrecen contraejemplos a la afirmación de que (de algún modo) la racionalidad es una facultad tal que, una vez que se la posee, puede extenderse hasta abarcar todos los conceptos, de manera que si se puede tener *algún* concepto, entonces se pueden tener *todos* los conceptos.

para nosotros; tendríamos una teoría explicativa del nexo causal en cuestión. Tendríamos el mismo tipo de comprensión que el de nuestras propias experiencias si tuviéramos la teoría psicofísica correcta de ellas. Pero entonces parece seguirse que la comprensión de la teoría que explica las experiencias *B* *conferiría* una comprensión de la naturaleza de esas experiencias, pues, ¿cómo podríamos comprender esa teoría sin comprender el concepto *B* que aparece en ella? ¿Cómo podríamos comprender la *naturaleza* de las experiencias *B* sin comprender el *carácter* de esas experiencias? La verdadera teoría psicofísica parecería ofrecer una vía para comprender la forma subjetiva de las experiencias del murciélago. Pero ahora nos enfrentamos a un dilema, un dilema que amenaza con convertirse en una *reductio*: o bien *podemos* comprender esa teoría, en cuyo caso la propiedad *B* se convierte en algo abierto para nosotros, o bien *no podemos* comprender esa teoría simplemente porque la propiedad *B* *no* se encuentra abierta para nosotros. Me parece aquí que la amenaza de *reductio* es persuasiva: nuestros conceptos de conciencia, simplemente *están* constreñidos inherentemente por nuestra propia forma de conciencia, de manera tal que cualquier teoría cuya comprensión requiera que trascendamos esos límites sería *ipso facto* una teoría inaccesible para nosotros. De manera similar, pienso que cualquier teoría que requiera que trascendamos la finitud de nuestras capacidades cognoscitivas sería *ipso facto* una teoría que no podríamos comprender —y ello a pesar de que podría necesitarse para explicar algo, según nos damos cuenta, necesita explicarse. No podemos simplemente estipular que nuestras capacidades para formar conceptos tienen una plasticidad y unos límites indefinidos sólo porque así tendrían que ser para permitirnos captar la verdad sobre el mundo. Debido a nuestra constitución, carecemos de la capacidad para formar conceptos que abarquen todos los tipos posibles de estado consciente, y esto obstruye el camino hacia una solución general del problema mente-cuerpo. E incluso si pudiéramos resolverlo para nuestro propio caso, no podríamos resolverlo para el caso de los murciélagos y los marcianos. Por decirlo así, *P* está demasiado cerca de las diferentes formas de subjetividad como para que sea accesible a todas ellas, dado que la forma

de subjetividad propia de cada quien restringe los conceptos relativos a la subjetividad.[12]

Sospecho que la mayoría de los que son optimistas acerca de una solución constructiva del problema mente-cuerpo preferirán colocar sus apuestas del lado del cerebro. Dirán que la neurociencia es el lugar donde hay que buscar la propiedad *P*. Mi pregunta es entonces si existe algún modo concebible en el que podamos llegar a introducir *P* en el curso de nuestras investigaciones empíricas sobre el cerebro. Ciertamente se han introducido conceptos nuevos en el esfuerzo por comprender el funcionamiento del cerebro: ¿no podría entonces aparecer *P* en desarrollos concebibles de esta forma de introducción? Hasta ahora, de hecho, los conceptos teóricos que atribuimos al cerebro parecen tan lejanos de la conciencia como cualquier propiedad física ordinaria, pero quizá podamos alcanzar *P* mediante una aplicación diligente de básicamente los mismos procedimientos: tal pensamiento resulta tentador. Deseo sugerir, por el contrario, que tales procedimientos están inherentemente cerrados con respecto a *P*. Pienso que la razón fundamental de ello es el papel de la *percepción* en la formación de nuestra comprensión del cerebro –la manera como nuestra percepción del cerebro limita los conceptos que podemos aplicarle. Sería difícil exagerar la importancia de la idea de que la propiedad de la conciencia misma (o de estados conscientes específicos) no es una propiedad observable o perceptible del cerebro. Puedes asomarte al interior de un cerebro consciente vivo, el tuyo

[12] Podría sugerirse que se tomara prestada la idea de Nagel de una "fenomenología objetiva" para salvar este escollo. En vez de representar las experiencias mediante descripciones subjetivas, deberíamos describirlas con términos enteramente objetivos, colocándolas así a nuestro alcance conceptual. Mi problema con esta posición es que, incluso suponiendo que pudiera haber semejante forma de descripción, no nos permitiría comprender cómo los aspectos subjetivos de la experiencia dependen del cerebro –lo que es realmente el problema que estamos tratando de resolver. De hecho, dudo que la noción de una fenomenología objetiva sea más coherente que la noción de una fisiología subjetiva. Ambas implican el intento de llenar el hueco psicofísico mediante una especie de estipulación. La lección aquí es que el hueco no puede ser llenado simplemente mediante la aplicación de conceptos tomados de un extremo a los objetos que pertenecen al otro extremo, y ello porque ninguno de esos tipos de conceptos podrá jamás llevar a cabo lo que se necesita.

o el de alguien más, y observar ahí una amplia variedad de propiedades instanciadas –su figura, color, textura, etc.–, pero no podemos *ver* por ello lo que el sujeto experimenta, el estado consciente mismo. Los estados conscientes simplemente no son objetos potenciales de la percepción: dependen del cerebro pero no pueden ser observados dirigiendo los sentidos hacia el cerebro. En otras palabras, la conciencia es nouménica con respecto a la percepción del cerebro.[13] Considero que esto es obvio. De modo que sabemos que *hay* propiedades del cerebro que están cerradas necesariamente a la percepción del cerebro; la pregunta ahora es si también *P* está cerrada a la percepción.

Mi argumento procederá del siguiente modo: en primer lugar argumentaré que *P* está de hecho perceptualmente cerrada; después completaré el argumento sosteniendo una clausura cognoscitiva completa insistiendo en que ninguna forma de *inferencia* a partir de lo percibido puede llevarnos a *P*. El argumento en favor de la clausura perceptual comienza con la idea de que nada de lo que podamos imaginar que percibimos en el cerebro podría convencernos alguna vez de que hemos localizado el nexo inteligible que buscamos. Independientemente de qué propiedad recóndita podamos observar instanciada en el cerebro, nos sentiremos siempre desconcertados respecto a cómo esa propiedad podría dar lugar a la conciencia. Invito al lector a tratar de concebir una propiedad perceptible del cerebro que pueda disipar el sentimiento de misterio que acompaña nuestra contemplación del eslabón mente-cerebro: no creo que sea capaz de hacerlo. Es como intentar concebir una propiedad perceptible de una roca que mostrara claramente que la roca es consciente. De hecho, pienso que la imposibilidad misma de realizar eso es lo que reside en la raíz del problema

[13] Deberíamos distinguir entre dos afirmaciones acerca de la imperceptibilidad de la conciencia: (i) la conciencia no es perceptible dirigiendo los sentidos hacia el cerebro; (ii) la conciencia no es perceptible dirijamos adonde dirijamos los sentidos, incluso aunque los dirijamos hacia la conducta que "expresa" los estados conscientes. Creo en ambas tesis, pero para lo que quiero sostener me basta (i). Estoy suponiendo, desde luego, que la percepción no puede estar cargada de teoría de una manera irrestricta; o, si lo puede estar, dichas cargas de teoría no pudieron haber sido derivadas originalmente simplemente viendo las cosas, o probándolas, o tocándolas, o...

mente-cuerpo tal como lo concebimos. Pero, ¿por qué es así? Pienso que básicamente se debe a que los sentidos están dirigidos hacia la representación de un mundo espacial; presentan esencialmente cosas en el espacio con propiedades espacialmente definidas. Pero son precisamente *tales* propiedades las que parecen ser inherentemente incapaces de resolver el problema mente-cuerpo: no podemos vincular la conciencia con el cerebro en virtud de las propiedades espaciales del cerebro. He ahí el cerebro: un objeto de la percepción, situado en el espacio y que contiene procesos espacialmente distribuidos; pero la conciencia no acepta una explicación en tales términos. La conciencia no parece estar constituida por procesos espaciales más pequeños; pero la percepción del cerebro parece estar limitada a revelar tales procesos.[14] Los sentidos responden a ciertos *tipos* de propiedades –las que están ligadas esencialmente con el espacio–, pero esas propiedades pertenecen al tipo incorrecto (a la *categoría* incorrecta) para constituir *P*. Kant estaba en lo cierto: la forma de nuestra sensibilidad externa es espacial; pero, si ello es así, entonces *P* será nouménico con respecto a los sentidos, ya que ninguna propiedad espacial proporcionará jamás una respuesta satisfactoria al problema mente-cuerpo. Simplemente no comprendemos la idea de que los estados conscientes puedan surgir de manera inteligible de configuraciones espaciales del tipo revelado por la percepción del mundo.

Supongo que esta tesis no parecerá terriblemente polémica. Después de todo, en general no esperamos que toda propiedad mencionada en nuestras teorías deba ser objeto potencial de la percepción humana: considérese la teoría cuántica y la cosmología. La apertura perceptual irrestricta es uno de los más claros dogmas del empirismo, y no hay ninguna razón poderosa para suponer que la propiedad que se necesita para explicar la relación mente-cerebro debiera ser en principio perceptible; podría

[14] Nagel discute la dificultad de pensar en los procesos conscientes en los términos espaciales aplicables al cerebro en Nagel 1986, pp. 50–51, pero no infiere mi conclusión desesperada. El caso es exactamente *distinto* de (digamos) la dependencia que tiene la liquidez en las propiedades de las moléculas, puesto que aquí sí pensamos en términos espaciales en ambos lados de la relación; por tanto, podemos simplemente emplear la idea de composición espacial.

ser esencialmente "teórica", un objeto de pensamiento y no de experiencia sensorial. Observar con mayor ahínco la naturaleza no es la única (o la mejor) manera de descubrir sus propiedades teóricamente significativas. La clausura perceptual no implica la clausura cognoscitiva, ya que tenemos a nuestra disposición el procedimiento de formación de hipótesis, en el cual se conceptualizan entes *in*observables.

Concuerdo fácilmente con esos sentimientos, pero pienso que hay razones para creer que ningún método coherente de introducción de conceptos nos *llevará* alguna vez a *P*. Ello se debe a que cierto principio de *homogeneidad* opera en nuestra introducción de conceptos teóricos sobre la base de la observación. Permítaseme señalar, en primer lugar, que la conciencia misma no podría ser introducida simplemente sobre la base de lo que observamos acerca del cerebro y sus efectos físicos. Si nuestros datos obtenidos mediante la percepción del cerebro no incluyen nada que introduzca estados conscientes, entonces las propiedades teóricas que necesitamos para explicar estos datos tampoco incluirán estados conscientes. La inferencia hacia la mejor explicación de datos puramente físicos nunca nos *conducirá* fuera del dominio de lo físico para forzarnos a introducir conceptos de conciencia.[15] Todo lo físico posee una explicación puramente física. Así, la propiedad de la conciencia está cognoscitivamente cerrada con respecto a la introducción de conceptos por medio de la inferencia hacia la mejor explicación de datos perceptuales sobre el cerebro.

Ahora la cuestión es si alguna vez podríamos alcanzar *P* por medio de este tipo de inferencia. Aquí debemos cuidarnos de no caer en una forma de emergentismo mágico con respecto a la formación de conceptos. Supóngase que probamos una teoría relativamente clara de cómo se forman los conceptos teóricos: los obtenemos mediante una especie de extensión analógica a partir de lo que observamos. Así, por ejemplo, obtenemos el concepto de molécula echando mano de nues-

[15] *Cfr.* Nagel 1979: "nunca será legítimo inferir, como una explicación teórica de fenómenos puramente físicos, una propiedad que incluye o implica la conciencia de su sujeto", "Panpsychism", p. 183.

tras representaciones preceptuales de objetos macroscópicos y concibiendo objetos de una escala más pequeña de la misma clase general. Este método parece funcionar bastante bien para objetos materiales inobservables, pero no será útil para obtener *P*, pues las extensiones analógicas de las entidades que observamos en el cerebro son tan inútiles como las entidades originales lo eran como soluciones al problema mente-cuerpo. Necesitaríamos un método que dejara atrás de una manera mucho más radical la base de propiedades observacionales. Sin embargo, me parece que ni siquiera una concepción más laxa de inferencia hacia la mejor explicación produciría el resultado deseado: no serviría más para introducir *P* que lo que sirve para introducir la propiedad misma de la conciencia. Para explicar los datos físicos observados necesitamos únicamente propiedades teóricas tales que se relacionen con esos datos, no la propiedad que explica la conciencia, que no aparece entre esos datos. Puesto que no necesitamos a la conciencia para explicar esos datos, no necesitamos la propiedad que explica a la conciencia. En nuestras explicaciones de esos datos nunca iremos más allá de los datos perceptuales como se requiere para conectarlos explicativamente con la conciencia. Ésta es, de hecho, la razón de por qué la conciencia parece ser teóricamente epifenoménica en la tarea de explicar los sucesos físicos. Ningún concepto que sea necesario para explicar el funcionamiento del mundo físico será suficiente para explicar cómo el mundo físico produce la conciencia. De tal modo que si *P* es perceptualmente nouménico, entonces será nouménico con respecto a las inferencias explicativas basadas en la percepción. En consecuencia, no pienso que se pueda obtener *P* sólo mediante estudios empíricos del cerebro. Sin embargo, el cerebro *posee* esa propiedad, como posee la propiedad de la conciencia. Sólo una idea mágica de cómo obtenemos conceptos nos podría llevar a pensar que podemos obtener *P* percibiendo primero el cerebro y preguntando después qué se necesita para explicar lo que percibimos.[16] (El

[16] Sin duda es notable el hecho de que los microprocesos que han sido descubiertos en el cerebro mediante los métodos comunes no parecen estar más cerca de la conciencia que las propiedades macroscópicas del cerebro abierto

problema mente-cuerpo nos induce a la magia en más de un sentido.)

Para aclarar la postura que quiero sostener, será útil compararla con otro enfoque para explicar la fuente de la perplejidad que sentimos respecto del nexo mente-cerebro. He argumentado que no podemos saber qué propiedad del cerebro es responsable de la conciencia, y por lo tanto el eslabón mente-cerebro nos parece ininteligible. No obstante, podría afirmarse, existe otra explicación de ese sentimiento de misterio irremediable que no requiere que postulemos propiedades que nuestras mentes no pueden representar. Este enfoque alternativo afirma que, incluso si *ahora* tuviésemos una comprensión de *P*, sentiríamos *aún* que hay algo misterioso en el nexo, debido a un rasgo epistemológico especial de la situación, a saber, nuestra familiaridad con el cerebro y nuestra familiaridad con la conciencia están necesariamente mediadas por facultades cognoscitivas distintas, la percepción y la introspección. Así, la facultad mediante la cual captamos uno de los términos de la relación es necesariamente distinta de la facultad mediante la cual captamos el otro término. En consecuencia, no es posible para nosotros utilizar sólo una de esas facultades para comprender la naturaleza del nexo psicofísico. Ninguna facultad individual nos permitirá nunca comprender el hecho de que la conciencia depende del cerebro en virtud de la propiedad *P*. Ni la percepción por sí sola ni la introspección por sí sola nos permitirán nunca ser testigos de esa dependencia. Y ésta, insistiría mi crítico, es la verdadera razón por la cual la relación nos resulta desconcertante: no podemos darle sentido en términos de lo que produce una sola facultad cognoscitiva. Por lo tanto, incluso si ahora poseyéramos conceptos para las propiedades del cerebro que explican la conciencia, seguiríamos experimentando un sentimiento residual de ininteligibilidad; supondríamos aún

a la inspección casual. Tampoco las características "holísticas" más abstractas del funcionamiento del cerebro parecen estar en el camino correcto para revelarnos la naturaleza de la conciencia. Entre más hondo sondea la ciencia en el cerebro, más parece alejarse de la conciencia. Así, el conocimiento cada vez mayor del cerebro destruye nuestras ilusiones acerca de los tipos de propiedades que podrían descubrirse siguiendo esa ruta. La teoría neurofisiológica avanzada sólo parece hacer más profundo el milagro.

que algo misterioso ocurre allí. La necesidad de cambiar de una facultad a la otra produce en nosotros una ilusión de inexplicabilidad. Podríamos de hecho tener la explicación en este mismo momento y estar al mismo tiempo bajo la ilusión de que no la tenemos. Así, el diagnóstico correcto es que deberíamos reconocer la peculiaridad de la situación epistemológica y dejar de tratar de entender el nexo psicofísico de la manera en que entendemos otros tipos de nexos. Simplemente nos *parece* que nunca podremos descubrir una propiedad que haga inteligible el nexo.

Pienso que el enfoque anterior merece ser considerado en serio, pero dudo que su diagnóstico de nuestro predicamento sea correcto. Es cierto que el nexo problemático es captado esencialmente por facultades distintas, de modo que nunca revelará sus secretos a una sola facultad; pero dudo que nuestro sentido intuitivo de inteligibilidad esté gobernado de una manera tan rígida por la "condición-de-una-sola-facultad". ¿Por qué los hechos sólo *deberían* parecernos inteligibles si podemos concebir que los aprehendemos mediante un (tipo de) facultad cognoscitiva? ¿Por qué no habremos de aceptar que podemos reconocer conexiones inteligibles entre conceptos (o propiedades) incluso cuando esos conceptos (o propiedades) son atribuidos necesariamente utilizando facultades distintas? ¿No es una actitud sospechosamente empirista insistir en que sólo podemos dar sentido a un nexo causal si podemos concebirlo como un objeto de una sola facultad de aprehensión? ¿Pensaríamos lo mismo en relación con un nexo que necesitara de los sentidos del tacto y de la vista para comprender cada término de la relación? Supóngase (*per impossibile*) que nos ofrecieran *P* en una charola, como un regalo de Dios: ¿menearíamos aún la cabeza y nos preguntaríamos cómo podría eso resolver el misterio, víctimas aún de la ilusión de misterio generado por la dualidad epistemológica en cuestión? No, no pienso que esta sugerencia sea suficiente para explicar la apariencia milagrosa del nexo: es mejor suponer que estamos impedidos permanentemente para formar un concepto de aquello que explica ese nexo.

¿Cuánta fuerza tiene la tesis por la que estoy abogando? Permítaseme distinguir entre afirmaciones *absolutas* y afirmaciones *relativas* de clausura cognoscitiva. Un problema está cerrado

cognoscitivamente de manera absoluta si ninguna mente posible puede resolverlo; un problema está relativamente cerrado si ciertos tipos de mente pueden resolverlo en principio mientras que otras mentes de otro tipo no pueden hacerlo. Podemos suponer, con toda seguridad, que la mayoría de los problemas están sólo cerrados de manera relativa: las mentes de los armadillos no pueden resolver problemas de aritmética elemental, pero las mentes humanas sí pueden hacerlo. ¿Diremos que el problema mente-cuerpo está sólo cerrado relativamente o que está cerrado de manera absoluta? Esto depende de qué aceptemos como una posible mente formadora de conceptos, lo cual no es un asunto fácil. Si aceptamos que las mentes forman sus conceptos del cerebro y la conciencia de manera totalmente independientes de la percepción y la introspección, entonces tiene cabida la idea de que existen mentes posibles para las cuales el problema mente-cuerpo puede resolverse, y de una manera fácil. Pero si suponemos que *toda* formación de conceptos está atada a la percepción y a la introspección, aunque sea laxamente, entonces *ninguna* mente será capaz de comprender cómo se relaciona con su propio cuerpo —el carácter insoluble del asunto será absoluto. Creo que podemos darle más o menos sentido al primer tipo de mente si aprovechamos nuestra propia facultad de razonamiento *a priori*. Por ejemplo, nuestros conceptos matemáticos no parecen estar atados ni a la percepción ni a la introspección, por lo que parece que hay una manera de formar conceptos que opera sin las restricciones que identifiqué anteriormente. La sugerencia podría ser entonces que una mente que formara todos sus conceptos de esta manera —incluyendo sus conceptos del cerebro y de la conciencia— se encontraría libre de las predisposiciones que nos impiden a *nosotros* alcanzar la teoría correcta de cómo se conectan el cerebro y la conciencia. Una mente con esas características tendría que ser capaz de pensar en el cerebro y la conciencia en formas que prescindan totalmente de todo lo perceptual y lo introspectible —en forma semejante a como nosotros ahora (al parecer) pensamos sobre los números. Esta mente concebiría el nexo psicofísico en términos totalmente *a priori*. Quizá ésta sea la manera como debamos concebir la mente de Dios, y la comprensión divina de la relación mente-cuerpo. De cualquier modo, se necesitará

algo bastante radical si hemos de idear una mente que pueda escapar de los tipos de clausura que hacen insoluble el problema para nosotros —si estoy en lo correcto en mi diagnóstico de nuestra dificultad. *Si* el problema es sólo insoluble de manera relativa, entonces el tipo de mente que podrá resolverlo será muy diferente de la nuestra y de los tipos de mente que podemos concebir (por supuesto, también aquí podría haber clausura cognoscitiva). Ciertamente, saber si el problema es absolutamente insoluble me parece cuando menos una pregunta abierta; no me sorprendería que lo fuera.[17]

Mi posición es a la vez pesimista y optimista. Es pesimista en relación con las perspectivas de alcanzar una solución constructiva para el problema mente-cuerpo, pero es optimista respecto a nuestras esperanzas de eliminar la perplejidad filosófica. La idea central aquí es que no pienso que debamos hacer lo primero para conseguir lo segundo. Ello depende de una comprensión más bien especial de aquello en lo que consiste el problema filosófico. Lo que deseo sugerir es que la naturaleza de la conexión psicofísica tiene una explicación completa y nada misteriosa en cierta ciencia, pero que esta ciencia es inaccesible para nosotros como una cuestión de principio. Llamemos *T* a esta teoría científica explicativa: *T* es tan natural, prosaica y ajena a los milagros como cualquier teoría de la naturaleza; describe el vínculo entre la conciencia y el cerebro de una manera que no es más notable (ni alarmante) que la manera en que ahora describimos el vínculo entre el hígado y la bilis.[18] De

[17] El tipo de limitación que he identificado no es, por lo tanto, del tipo que podría ser superado simplemente mediante un gran incremento de inteligencia general. Independientemente de cuánto puedan crecer los lóbulos frontales de nuestros descendientes biológicos, seguirán estando confundidos por el problema mente-cuerpo mientras sigan formando sus conceptos (empíricos) sobre la base de la percepción y la introspección.

[18] O bien, no más milagrosa que la teoría de la evolución. El creacionismo es una respuesta comprensible al problema teórico planteado por la existencia de organismos complejos; por fortuna, hoy tenemos una teoría que hace innecesaria esta respuesta, y mina así el teísmo requerido por la tesis creacionista. En el caso de la conciencia, la apariencia milagrosa del fenómeno también podría tentarnos a tomar la vía "creacionista", en la que se necesita a Dios para que lleve a cabo la alquimia necesaria para transformar la materia en experiencia. Así, el problema mente-cuerpo podría, en términos similares,

acuerdo con *T*, no ocurre nada misterioso en el mundo cuando un suceso en mi corteza visual causa que yo tenga una experiencia del amarillo –por mucho que *nos* parezca que sí ocurre. En otras palabras, no existe ninguna dificultad intrínseca conceptual o metafísica respecto a cómo la conciencia depende del cerebro. No es que la ciencia correcta esté obligada a postular milagros *de re*; se trata más bien de que la ciencia correcta está en la parte del mundo que es oscura para nosotros. Confundimos nuestras propias limitaciones cognoscitivas con un carácter misterioso objetivo. Somos como una mente humana tratando de comprender el mundo físico, o una criatura que carece de conceptos espaciales tratando de entender la posibilidad del movimiento. Esto elimina el problema filosófico porque nos asegura que las entidades *mismas* no plantean ninguna dificultad filosófica inherente. El caso es distinto, por ejemplo, del problema de cómo el mundo abstracto de los números puede ser relacionado inteligiblemente con el mundo de los sujetos cognoscentes concretos: en este caso el misterio parece ser una cualidad intrínseca de las entidades, no un simple producto de las limitaciones o tendencias cognoscitivas al tratar de entender la relación.[19] No sería plausible sugerir que existe una ciencia, cuyos conceptos teóricos no podemos comprender, que resuelve de una manera completa cualquier sensación de misterio que circunda la pregunta de cómo lo abstracto se convierte en un objeto de conocimiento para nosotros. El eliminativismo pare-

servir para probar la existencia de Dios (no hay milagros sin un hacedor de milagros). Pienso que no podemos refutar este argumento del modo en que podemos refutar el argumento creacionista original, a saber, construyendo efectivamente una teoría explicativa no milagrosa; pero sí podemos refutarlo sosteniendo que tal teoría naturalista debe *existir*. (Evitar dar por supuesto el teísmo es una condición de adecuación para cualquier intento de explicación de la relación mente-cuerpo.)

[19] Véase Benacerraf (1973) para una formulación de este problema acerca de las entidades abstractas. Otro problema que me parece que difiere del problema mente-cuerpo es el de la voluntad libre. No creo que haya una propiedad incognoscible *Q* que reconcilie la voluntad libre con el determinismo (o el indeterminismo); más bien el concepto de voluntad libre contiene incoherencias internas –a diferencia del concepto de conciencia. Por ello es mucho más razonable ser un eliminativista en el problema de la voluntad libre que en el problema de la conciencia.

ce entonces, en este caso, una opción real. El problema *filosófico* acerca de la conciencia y el cerebro surge de la sensación de que estamos obligados a aceptar que en la naturaleza hay milagros —como si la lámpara de simple metal del cerebro pudiera realmente dotar de existencia al genio de la conciencia. Pero no tenemos que aceptar eso: podemos permanecer tranquilos en el conocimiento de que alguna propiedad (incognoscible) del cerebro hace que todo funcione correctamente. Lo que crea la perplejidad filosófica es el supuesto de que el problema debe ser de algún modo científico, pero que cualquier ciencia que *nosotros* podamos construir representará las cosas como si fueran totalmente milagrosas. Y la solución es reconocer que el sentido milagroso proviene de nosotros y no del mundo. En realidad, no hay nada misterioso en cómo genera el cerebro la conciencia. No hay un problema *metafísico*.[20]

Hasta ahora esa tesis deflacionaria ha sido justificada con un naturalismo general y ciertas consideraciones sobre la clausura cognoscitiva y las ilusiones a las que puede dar origen. Ahora deseo introducir algunas razones para pensar que la conciencia es, en realidad, un hecho natural más bien simple; en términos objetivos, la conciencia no es algo muy especial. A estas alturas deberíamos estar ya acostumbrados a la idea de que nuestro propio sentido de la dificultad es un guía falible para identificar la complejidad objetiva: lo que es difícil de comprender para nosotros podría no ser muy sofisticado de suyo. La textura de nuestro pensamiento no es un espejo de los hechos de la naturaleza.[21] En particular, pudiera ser que el alcance de nuestra

[20] Una manera de poner a prueba la adecuación de una solución propuesta al problema mente-cuerpo es averiguar si mitiga la presión ejercida hacia el eliminativismo. Si los datos sólo pueden ser explicados postulando un milagro (*i.e.*, si no se explican), entonces debemos rechazar los datos —éste es el principio que se encuentra detrás del impulso de negar que los estados conscientes existan. Mi propuesta pasa esta prueba porque nos permite evitar la postulación de milagros; interpreta el carácter misterioso en términos meramente epistémicos, aunque en un sentido profundo. Las soluciones constructivas no son la única manera de mitigar la presión.

[21] Chomsky sugiere que las mismas facultades de la mente que hacen que seamos buenos para realizar ciertas tareas cognoscitivas pueden hacernos malos para otras; véase Chomsky 1975, pp. 155–156. Me parece posible que lo que nos hace buenos para la ciencia del mundo puramente físico es lo que nos

comprensión de los hechos acerca de la mente no sea conmensurable con alguna manera objetiva de calcular su complejidad intrínseca: podríamos ser buenos para entender la mente en algunos de sus aspectos pero estar perdidos con respecto a otros, de una manera que va más allá de las diferencias objetivas acerca de lo que los aspectos entrañan. Así, somos expertos en comprender la acción en términos de la psicología popular de creencias y deseos, y no parecemos estar totalmente perdidos cuando se trata de construir teorías del lenguaje; pero nuestra comprensión de cómo se desarrolla la conciencia a partir de la organización de la materia simplemente no existe. Sin embargo, piénsese ahora en esos múltiples aspectos de la mente desde el punto de vista de la biología evolucionista. Seguramente el lenguaje y las actitudes proposicionales son logros evolutivos más complejos y avanzados que la simple posesión de la conciencia por parte de un organismo físico. Por lo tanto, parece que somos mejores para entender algunos de los aspectos más complejos de la mente que para entender otros más simples. La conciencia surge en una etapa temprana de la historia evolutiva y podemos encontrarla a todo lo largo del reino animal. En ciertos aspectos parece que la ingeniería biológica que se requiere para la conciencia es menos sofisticada que la que se necesita para ciertos tipos de conducta motriz compleja. No obstante, podemos llegar a entender lo segundo sin entender en absoluto lo primero. Hablando en términos comparativos, los estados conscientes parecen ser bastante primitivos biológicamente. De manera que la teoría *T* que explica la aparición de la conciencia en el mundo físico es muy probablemente menos compleja objetivamente (según algún estándar) que una gama de otras teorías que no desafían nuestro intelecto. Si tan sólo pudiéramos conocer el mecanismo psicofísico nos sorprenderíamos por su sencillez, su carácter absolutamente natural. En el manual que Dios con-

desvía de la posibilidad de desarrollar una ciencia de la conciencia. Nuestras facultades hacen que nos inclinemos a comprender la materia en movimiento, pero es justamente esta clase de comprensión la que resulta inaplicable al problema mente-cuerpo. Quizás, entonces, el precio de ser buenos para entender la materia es que no podamos entender la mente. Ciertamente nuestra notable tendencia a pensar todo en términos espaciales no nos ayuda a entender la mente.

sultó cuando hizo la Tierra y a todas las bestias que habitaron en ella, el capítulo sobre cómo fabricar la conciencia a partir de la materia aparece más o menos cerca del principio, mucho antes que los capítulos realmente difíciles sobre la reproducción de los mamíferos y el habla. No es el *tamaño* del problema sino su *tipo* lo que hace tan difícil para nosotros el problema mente-cuerpo. Esta reflexión nos debería hacer receptivos a la idea de que es algo acerca de los procesos de nuestro pensamiento lo que nos impide alcanzar una ciencia que relacione a la conciencia con su base física: el enemigo está en casa.[22]

La postura a la que he llegado tiene implicaciones para una maraña de intuiciones que es natural poseer con respecto a la relación mente-cuerpo. Por un lado, hay intuiciones, formuladas desde Descartes hasta Kripke, según las cuales la relación entre los estados conscientes y los estados del cuerpo es fundamentalmente contingente.[23] Fácilmente podría parecernos que no hay una necesidad implicada en la dependencia de la mente con respecto al cerebro. Pero, por otro lado, parece absurdo tratar de disociar totalmente los dos términos, dejar que la mente subsista completamente libre del cuerpo. La incorporeidad es, en el mejor de los casos, una posibilidad dudosa, y muchos han creído innegable algún tipo de superveniencia necesaria de lo mental en lo físico. No es mi intención aquí determinar un vencedor en esta ya larga disputa; simplemente deseo ofrecer un diagnóstico de lo que ocurre cuando uno se encuentra abrumado por esta ráfaga de intuiciones en conflicto. La razón por la cual sentimos la fuerza de la contingencia, tratando de liberar a la conciencia de su amarre físico, podría ser que no

[22] He tomado esta frase de Fodor 1983, p. 121. El contraste que se pretende resaltar es con tipos de clausura cognoscitiva que surgen debido a factores exógenos —como, por ejemplo, en astronomía. Nuestro problema con *P* no es que esté muy distante, o que sea demasiado pequeño o demasiado grande o demasiado complejo; más bien la estructura misma de nuestro aparato para formar conceptos nos desvía de *P*.

[23] Saul Kripke 1980. Desde luego, Descartes argumentó explícitamente a partir de (lo que él consideró que eran) las naturalezas esenciales de la mente y el cuerpo para inferir la contingencia de su conexión. Si abandonamos el supuesto de que conocemos esas naturalezas, entonces el agnosticismo respecto a la modalidad de la conexión parece ser la conclusión indicada.

entendemos y no podemos entender la naturaleza de la propiedad que los vincula de manera inteligible. El cerebro tiene propiedades físicas que podemos comprender, y ciertas variaciones de ellas se correlacionan con cambios en la conciencia, pero no podemos correr el velo que oculta la manera en que se conectan. Al no comprender la naturaleza de la conexión, nos parece como algo profundamente contingente; no podemos hacer inteligible para nosotros mismos la afirmación de una conexión necesaria. *Podría* haber entonces una verdadera conexión necesaria; simplemente sucede que siempre nos parecerá algo curiosamente primario y oscuro. Puede ser entonces que, como defensores de la contingencia intrínseca, seamos las inocentes víctimas de nuestra propia ceguera cognoscitiva. Por otro lado, difícilmente estamos en condiciones de afirmar que *hay* una conexión necesaria entre las propiedades que podemos comprender del cerebro y los estados de la conciencia, pues estamos en la ignorancia plena (y de una manera irremediable) con respecto al carácter de la conexión. Por lo que sabemos, la conexión podría ser contingente, como el acceso a *P* revelaría si tal acceso fuese posible. El nexo entre la conciencia y la propiedad *P* no es, sin duda, contingente –virtualmente por definición– pero no estamos en condiciones de decir exactamente cómo se relaciona *P* con las propiedades "ordinarias" del cerebro. Podría ser necesaria o contingente. Así resulta que tendemos a vacilar entre la contingencia y la necesidad, pues carecemos de los recursos conceptuales para decidir el asunto –o para comprender la respuesta que tendemos a ofrecer. La conclusión indicada parece ser que nunca podremos saber realmente si la incorporeidad es metafísicamente posible, o si la superveniencia necesaria es el caso, o si podría suceder la inversión del espectro; pues todas estas respuestas implican supuestos acerca de las conexiones modales entre las propiedades de la conciencia y las propiedades ordinarias del cuerpo y el cerebro que podemos conceptualizar, y la verdadera naturaleza de estas conexiones es inaccesible para nosotros. Quizá *P* hace que la relación entre los disparos de las fibras C y el dolor sea necesaria, o quizá no: simplemente no estamos capacitados para saberlo. Somos como una mente humeana preguntándose si el vínculo observado entre la temperatura de un gas y su presión

(en un volumen constante) es necesario o contingente. Para saber la respuesta se necesita comprender la teoría atómica (o molecular), y una mente humeana simplemente no está a la altura para entender tales requisitos teóricos. De manera similar, somos constitutivamente ignorantes precisamente ahí donde está la respuesta.

Supongo que a muchos lectores la tesis principal de este artículo les parecerá totalmente increíble e incluso ridícula. Permítaseme observar que simpatizo con ellos: la tesis no se digiere fácilmente. Pero señalaría lo siguiente: si la tesis *es* de hecho verdadera, de cualquier modo nos seguirá pareciendo difícil de creer. Pues la idea de una propiedad (o conjunto de propiedades) explicativa que sea nouménica para nosotros, pero que sea esencial para la solución (constructiva) de un problema al que nos enfrentamos, puede resultar ofensiva para un tipo de idealismo natural que tiende a dominar nuestro pensamiento. Nos parece oneroso concebir la existencia de una propiedad real, bajo nuestras narices, por decirlo así, y que estamos construidos para no comprender —una propiedad responsable de fenómenos que observamos de la manera más directa posible. Esta clase de realismo, que acerca tanto la clausura cognoscitiva a casa, puede parecer tanto una afrenta a nuestro intelecto como algo imposible de evadir. Tratamos de pensar en esta propiedad impensable y, de manera comprensible, fracasamos en nuestros esfuerzos; de modo que nos apresuramos a inferir que el supuesto mismo de tal propiedad carece de sentido. Así, parece difícil mantener enfocado el realismo de la clase que estoy suponiendo, y cualquier teoría filosófica que dependa de él parecerá también descansar sobre algo sistemáticamente elusivo.[24] Sin embargo, mi respuesta a tales dudas no concede

[24] Éste es el tipo de realismo defendido por Nagel en el cap. VI de Nagel 1986: ser no es ser concebible por nosotros. Pienso que el problema mente-cuerpo ofrece una demostración de que *hay* tales propiedades que trascienden nuestros conceptos —y no simplemente de que *podría* haberlas. También pienso que el realismo de esta clase debería aceptarse precisamente porque ayuda a resolver el problema mente-cuerpo; es una tesis metafísica que concentra su atractivo en el hecho de que sale adelante con un problema que de otro modo parecería ser un caso perdido. No hay por tanto nada "epifenoménico" respecto a tal realismo radical: la existencia de una realidad que no pode-

nada: los límites de nuestras mentes simplemente no son los límites de la realidad. Insistir en que la realidad se encuentra limitada por lo que la mente humana puede concebir es una actitud deplorablemente antropocéntrica. Necesitamos cultivar una concepción de la realidad (una metafísica) que la haga verdaderamente independiente de nuestras capacidades cognoscitivas dadas, una concepción que incluya esas capacidades como una parte propia. Sucede que, en el caso del problema mente-cuerpo, el trozo de realidad que evade sistemáticamente nuestro alcance cognoscitivo es un aspecto de nuestra propia naturaleza. De hecho, es un aspecto que hace posible que tengamos mente y que podamos pensar en cómo se relaciona con nuestro cuerpo. Resulta que este particular trecho trascendente de realidad reside dentro de nuestras propias cabezas. Así, queda necesariamente oculto para nosotros un hecho profundo acerca de nuestra propia naturaleza en tanto que una forma de conciencia encarnada. Pero no hay nada que sea inherentemente misterioso o extravagante acerca de esta encarnación. Somos mucho más simples de lo que parecemos. Nuestra rareza reside en el ojo de quien observa.

La respuesta a la pregunta de mi título es, por lo tanto, "No y Sí".

[*Traducción de Hector Islas*]

BIBLIOGRAFÍA

Benacerraf, P., 1973, "Mathematical Truth", *Journal of Philosophy*, vol. 70, no. 19, pp. 661–679.

Chomsky, N., 1975, *Reflections on Language*, Patheon Books, Nueva York. [Versión en castellano: *Reflexiones sobre el lenguaje*, trad. J. Argente y J. Nadal, Ariel, Barcelona, 1986.]

Fodor, J., 1983, *The Modularity of Mind*, MIT Press, Cambridge, Mass. [Versión en castellano: *La modularidad de la mente*, trad. J.M. Igoa, Morata, Madrid, 1986.]

mos conocer puede tener, de cualquier modo, importancia intelectual para nosotros.

Kripke, S., 1980, *Naming and Necessity*, Blackwell, Oxford. [Versión en castellano: *El nombrar y la necesidad*, trad. M.M. Valdés, Instituto de Investigaciones Filosóficas-UNAM, México, 1995.]

Locke, J., (1689), *Essay Concerning Human Understanding*, Clarendon Press, Oxford, 1975. [Versión en castellano: *Ensayo sobre el entendimiento humano*, trad. L. Rodríguez Aranda, Aguilar, Buenos Aires, 1977.]

Nagel, T., 1986, *The View From Nowhere*, Oxford University Press, Oxford. [Versión en castellano: *Una visión de ningún lugar*, trad. Fondo de Cultura Económica, México, 1996.]

——, 1979, *Mortal Questions*, Cambridge University Press, Cambridge.

Strawson, P.F., 1959, *Individuals*, Methuen, Londres. [Versión en castellano: *Individuos*, trad. A. García Suárez y L.M. Valdés Villanueva, Taurus, Madrid, 1989.]

QUALIA EPIFENOMÉNICOS*

FRANK JACKSON

Es innegable que las ciencias físicas, químicas y biológicas han proporcionado mucha información acerca del mundo en que vivimos y acerca de nosotros mismos. Utilizaré la etiqueta "información física" para esta clase de información y también para la información que automáticamente la acompaña. Por ejemplo, si un científico médico me dice lo suficiente acerca de los procesos que ocurren en mi sistema nervioso, y acerca de cómo se relacionan con los acontecimientos del mundo que me rodea, con lo que ha sucedido en el pasado y lo que es probable que suceda en el futuro, con lo que les sucede a otros organismos similares y distintos, y cosas semejantes, él o ella me dice –si soy lo suficientemente inteligente para ensamblarlo todo apropiadamente– lo que se suele llamar el papel funcional de esos estados en mí (y en los organismos en general en casos similares). A esta información y a otras del mismo tipo también las calificaré de "físicas".

No quiero decir que estas observaciones esquemáticas constituyan una definición de la información "física", ni de las nociones correlativas de propiedad física, proceso físico, etc., sino que sirven para indicar lo que aquí estoy pensando. Es bien conocido que existen problemas para dar una definición precisa de estas nociones y, por consiguiente, de la tesis del fisicalismo

* Estoy en deuda con Robert Pargetter por varios comentarios y, pese a su desacuerdo, con el §IV de Meehl 1966.

de que toda información (correcta) es información física.[1] Pero —a diferencia de algunos— me parece que esta cuestión de la definición trasciende los problemas centrales que quiero discutir en este artículo.

Soy lo que algunos llaman un "fanático de los *qualia*". Creo que hay ciertos rasgos de las sensaciones corporales en especial, pero también de ciertas experiencias perceptuales, que no está incluida en ninguna cantidad de información física. Díganme todo lo físico que se puede decir acerca de lo que ocurre en un cerebro vivo, los tipos de estados, sus papeles funcionales, sus relaciones con lo que ocurre en otros momentos y en otros cerebros, etc., etc., y aunque sea todo lo inteligente que se tiene que ser para unir toda la información, no se me habrá dicho nada acerca de lo doloroso de los dolores, del picor de las picazones, de las punzadas de los celos, o acerca de la experiencia característica de probar un limón, oler una rosa, oír un ruido fuerte o ver el cielo.

Existen muchos fanáticos de los *qualia*, y algunos de ellos dicen que su rechazo al fisicalismo es una intuición sin argumentos.[2] Creo que están siendo injustos consigo mismos. Tienen el siguiente argumento. Nada de lo que se pueda decir en el ámbito de lo físico capta el olor de una rosa, por ejemplo. Por lo tanto, el fisicalismo es falso. Según nosotros, éste es un argumento perfectamente bueno. Obviamente, no viene al caso cuestionar su validez, y la premisa es, desde un punto de vista intuitivo, obviamente verdadera para ellos y para mí.

Sin embargo, debo admitir que es débil desde un punto de vista polémico. Por desgracia para nosotros, hay muchos a quienes la premisa no les parece intuitivamente obvia. La tarea consiste entonces en ofrecer un argumento cuyas premisas sean obvias para todos, o al menos para el mayor número posible. Es esto lo que trato de hacer en la sección 1, con lo que denomino "el argumento del conocimiento". En la sección 2, contrasto el argumento del conocimiento con el argumento modal y en la 3 con el argumento "cómo es ser". En la sección 4, abordo

[1] Véase, por ejemplo, Mellor 1973; y Cornman 1971.

[2] Particularmente en discusiones, pero véase, por ejemplo, Campbell 1976, p. 67.

la cuestión del papel causal de los *qualia*. El principal factor que impide que la gente admita los *qualia* es la creencia de que debería atribuírseles un papel causal respecto del mundo físico y, en especial, del cerebro;[3] y es difícil hacerlo sin sonar como alguien que cree en las hadas. En la sección 4, intento voltear esta objeción arguyendo que la idea de que los *qualia* son epifenoménicos es del todo posible.

1. *El argumento del conocimiento en favor de los* qualia

La gente varía considerablemente en su capacidad para discriminar colores. Supóngase que en un experimento para catalogar esta variación se descubre a Alfredo. Alfredo tiene la mejor visión del color que cualquiera de los que se tiene registro; ha hecho todas las discriminaciones nunca antes hechas, y además hace otra que el resto ni siquiera podemos empezar a hacer. Si se le muestra un montón de tomates maduros, los clasificará en dos grupos aproximadamente iguales y lo hará con absoluta consistencia. Es decir, si le vendan los ojos, se mezclan los tomates y luego se le retira la venda y se le pide que los clasifique de nuevo, los clasificará exactamente en los mismos dos grupos.

Le preguntamos a Alfredo cómo lo hace. Él explica que no todos los tomates maduros se ven del mismo color para él, y que de hecho esto es cierto de una gran cantidad de objetos que nosotros clasificamos juntos como rojos. Él ve dos colores donde nosotros vemos uno, y en consecuencia ha desarrollado para su propio uso dos palabras "rojo$_1$" y "rojo$_2$" a fin de marcar la diferencia. Tal vez nos diga que a menudo ha tratado de enseñar a sus amigos la diferencia entre "rojo$_1$" y "rojo$_2$", pero no ha tenido éxito y ha llegado a la conclusión de que el resto del mundo es daltónico respecto a los colores rojo$_1$ y rojo$_2$ —o quizá haya tenido un éxito parcial con sus hijos, pero no importa. De cualquier modo, nos explica que sería bastante equivocado pensar que porque "rojo" aparece tanto en "rojo$_1$" como en "rojo$_2$", ambos colores son tonos de un mismo color. Él sólo usa el término común "rojo" para ajustarse con mayor facilidad a nuestro uso limitado. Para él, el rojo$_1$ y el rojo$_2$ son tan diferentes entre

[3] Véase, por ejemplo, Dennett 1978.

sí y del resto de los colores, como lo es el amarillo del azul. Y su comportamiento discriminativo lo confirma: distingue los tomates $rojo_1$ de los $rojo_2$ con la mayor facilidad en una amplia variedad de circunstancias de observación. Además, una investigación de la base fisiológica de la excepcional capacidad de Alfredo revela que su sistema óptico es capaz de separar dos grupos de longitudes de onda en el espectro del rojo con tanta nitidez como los demás somos capaces de distinguir el amarillo del azul.[4]

Creo que deberíamos admitir que Alfredo puede ver, realmente ver, al menos un color más que nosotros; el $rojo_1$ es distinto del $rojo_2$. Somos, para Alfredo, tan daltónicos como lo son para nosotros quienes no distinguen entre el rojo y el verde. El cuento de H.G. Wells "El país de los ciegos" trata de una persona que ve en una comunidad donde todos son ciegos.[5] Esta persona nunca logra convencerlos de que puede ver, de que tiene un sentido más. Los ciegos ridiculizan este sentido tachándolo de bastante inconcebible, y toman su capacidad para evitar caer en zanjas, para ganar peleas, etc., como esa capacidad y nada más. Estaríamos cometiendo ese mismo error si nos rehusáramos a aceptar que Alfredo puede ver un color más que nosotros.

¿Qué clase de experiencia tiene Alfredo cuando ve el $rojo_1$ y el $rojo_2$? ¿Cómo es o cómo son los nuevos colores? Cuánto daríamos por saberlo, pero no lo sabemos; y parece que no hay ninguna cantidad de información física acerca del cerebro y del sistema óptico de Alfredo que nos lo pueda decir. Tal vez descubrimos que los conos de Alfredo responden diferencialmente a ciertas ondas luminosas en la sección del rojo del espectro que en nosotros no hacen diferencia alguna (o quizá él tenga un cono más), y que eso hace que Alfredo tenga una gama más amplia de los estados mentales responsables del comportamiento visual discriminatorio. Pero nada de esto nos dice lo que en realidad queremos saber acerca de su experiencia de los colores. Hay algo acerca de ella que no sabemos. Pero sí sabemos,

[4] Si así se prefiere, póngase ésta, y las simplificaciones similares que aparecen más adelante, en términos de la teoría de Land. Véase, por ejemplo, Land 1959.

[5] Wells, s.f.

supongámoslo, todo lo relacionado con el cuerpo de Alfredo, con su comportamiento y sus disposiciones conductuales, con su fisiología interna, y todo lo que tiene que ver con su historia y su relación con los demás que puede figurar en las explicaciones físicas de las personas. Tenemos toda la información física. Por consiguiente, saber todo esto *no* equivale a saber todo acerca de Alfredo. De aquí se sigue que el fisicalismo está dejando algo fuera.

Para reforzar esta conclusión, imaginemos que como resultado de nuestras investigaciones de los procesos internos de Alfredo, descubrimos cómo hacer que la fisiología de cada quien se asemeje a la de Alfredo en los aspectos relevantes; o tal vez Alfredo dona su cuerpo a la ciencia y al morir logramos trasplantar su sistema óptico a alguien más –tampoco en este caso importan mucho los detalles. Lo importante es que ese suceso despertaría un interés enorme. La gente diría: "Al fin vamos a saber cómo es ver el otro color, al fin vamos a saber en qué se diferenciaba Alfredo de nosotros tal como él luchó por explicárnoslo durante tanto tiempo." Entonces, no puede ser que siempre hayamos sabido todo acerca de Alfredo. Pero, *ex* hipótesis, sí supimos siempre acerca de Alfredo todo lo que figura en el esquema fisicalista; por lo tanto, el esquema fisicalista deja algo fuera.

Plantcémoslo de la siguiente manera. *Después* de la operación, sabremos *más* acerca de Alfredo y en especial acerca de sus experiencias de los colores. Pero antes tuvimos toda la información física que podríamos desear acerca de su cuerpo y su cerebro, y de hecho todo lo que siempre ha figurado en las explicaciones fisicalistas de la mente y la conciencia. Por lo tanto, hay más por saber que todo eso. Por lo tanto, el fisicalismo es incompleto.

Alfredo y el (los) nuevo(s) color(es) son, desde luego, en esencia, recursos retóricos. El mismo argumento podría hacerse refiriéndonos a gente normal y a colores conocidos. María es una brillante científica que, por alguna razón, ha sido forzada a investigar el mundo desde una habitación en blanco y negro, mediante un monitor de televisión en blanco y negro. Se especializa en neurofisiología de la visión y adquiere, supongámoslo, toda la información física que hay que obtener

acerca de lo que pasa cuando vemos tomates maduros, o el cielo, y usamos términos como "rojo", "azul", y demás. Descubre, por ejemplo, precisamente qué combinaciones de longitudes de onda del cielo estimulan la retina, y exactamente cómo esto produce, *mediante* el sistema nervioso central, la contracción de las cuerdas vocales y la expulsión de aire de los pulmones que dan por resultado que se pronuncie la oración "El cielo es azul". (Difícilmente se puede negar que en principio es posible obtener toda esta información física por medio del televisor en blanco y negro; de otro modo, *por necesidad*, la Open University británica tendría que usar televisores a color.)

¿Qué pasará cuando se libere a María de su habitación en blanco y negro o cuando se le dé un monitor de televisión a color? ¿*Aprenderá* algo o no? Parece perfectamente obvio que aprenderá algo acerca del mundo y de la experiencia visual que tenemos de él. Pero entonces es ineludible que su conocimiento previo era incompleto. Sin embargo, tenía *toda* la información física. *Ergo*, hay más conocimiento por obtener que eso, y el fisicalismo es falso.

Evidentemente, se podría presentar un argumento del conocimiento del mismo tipo para el gusto, el oído, las sensaciones corporales y, en general, para los diversos estados mentales de los que se dice tienen (en los muy diversos términos que se han usado) sensaciones crudas, rasgos fenoménicos o *qualia*. La conclusión en cada caso es que los *qualia* quedan fuera de la explicación fisicalista. Y la fuerza polémica del argumento del conocimiento consiste en que es muy difícil negar la afirmación central de que se puede tener toda la información física sin tener toda la información que podría tenerse.

2. *El argumento modal*

Cuando hablo del argumento modal me refiero a un argumento del siguiente tipo.[6] Los escépticos acerca de otras mentes no están cometiendo ningún error de lógica deductiva, no importa en qué otros aspectos esté equivocada su posición. Ninguna cantidad de información física acerca de otro *implica lógicamente*

[6] Véanse, por ejemplo, Campbell 1970; y Kirk 1974.

que él o ella sea consciente o que sienta algo en absoluto. En consecuencia, hay un mundo posible con organismos exactamente como nosotros en todos los aspectos físicos (y recuérdese que esto incluye estados funcionales, historia física, *et al.*), pero que difieren de nosotros profundamente en que no tienen ningún tipo de vida mental consciente. Pero luego, ¿qué es lo que nosotros tenemos y de lo cual ellos carecen? No es nada físico *ex* hipótesis. En todos los aspectos físicos, ellos y nosotros somos exactamente iguales. En consecuencia, nosotros somos algo más que lo puramente físico. Luego, el fisicalismo es falso.[7]

A veces se objeta que el argumento modal interpreta incorrectamente el fisicalismo sobre la base de que esa doctrina se presenta como una verdad *contingente*.[8] Pero decir esto sólo equivale a decir que los fisicalistas restringen su afirmación a algunos mundos posibles, entre los que se incluye especialmente el nuestro; y el argumento modal se dirige sólo en contra de esta afirmación menor. Si nosotros en *nuestro* mundo, ya no se diga seres de cualquier otro mundo, tenemos rasgos adicionales a los de nuestras réplicas físicas de otros mundos posibles, entonces tenemos rasgos no físicos o *qualia*.

El problema con el argumento modal es más bien que se apoya en una intuición modal discutible. Discutible porque se discute. Algunos niegan sinceramente que pueda haber réplicas físicas de nosotros en otros mundos posibles que, no obstante, carezcan de conciencia. Más aún, al menos una persona que alguna vez tuvo la intuición ahora tiene sus dudas.[9]

Contar el número de seguidores puede parecer un modo deficiente de enfocar la discusión del argumento modal. Pero a menudo no podemos hacer otra cosa cuando se trata de intuiciones modales, y recuérdese que nuestro objetivo inicial era encontrar el argumento con mayor utilidad polémica.

[7] He presentado el argumento usando una relación entre mundos y no del modo más usual de intramundo con el fin de evitar complicaciones innecesarias que tienen que ver con superveniencia, anomalías causales y similares.

[8] Véanse, por ejemplo, Lycan 1979, p. 280; y Locke 1976.

[9] Véase Kirk 1979. Véanse también los argumentos en contra de la intuición modal en, por ejemplo, Shoemaker 1975 [incluido en esta antología, pp. 183–212].

Desde luego, *qua* protagonistas del argumento del conocimiento bien podemos aceptar la intuición modal en cuestión; pero ésta sería una *consecuencia* de que ya teníamos un argumento para la conclusión de que los *qualia* quedan fuera de la explicación fisicalista, no nuestra base para esa conclusión. Además, el asunto se complica con la posibilidad de que la conexión entre lo físico y los *qualia* sea como la que a veces se considera que se da entre cualidades estéticas y cualidades naturales. Dos mundos posibles que coincidan en todos los aspectos "naturales" (incluso en las experiencias de las criaturas sensibles) tienen que coincidir también en todas las cualidades estéticas, pero se sostiene de un modo plausible que las cualidades estéticas no se pueden reducir a las naturales.

3. *El argumento "cómo es ser"*

En "¿Cómo es ser un murciélago?", Thomas Nagel argumenta que ninguna cantidad de información física puede decirnos cómo es ser murciélago, y que de hecho nosotros, los seres humanos, no podemos imaginarnos cómo es ser un murciélago.[10] Su razón es que cómo es ser esto sólo se puede entender desde el punto de vista de un murciélago, que no es nuestro punto de vista y que tampoco puede captarse en términos físicos que son, en esencia, términos que se pueden comprender de igual modo desde muchos puntos de vista.

Es importante distinguir entre éste y el argumento del conocimiento. Cuando me quejé de que todo el conocimiento físico acerca de Alfredo no era suficiente para decirnos cómo era su especial experiencia del color, mi queja no se dirigía a que no descubriríamos cómo era *ser* Alfredo. Me quejaba más bien de que hay algo *acerca de* su experiencia, una de sus propiedades, respecto del cual nos quedamos en la ignorancia. Y si

[10] Nagel 1974, pp. 435–450 [incluido en esta antología, pp. 45–63]. Hay que decir dos cosas acerca de este artículo. Una es que, pese a las diferencias que expondré, le debo mucho. La otra es que el énfasis cambia a lo largo del artículo y, hacia el final, Nagel ya no objeta tanto al fisicalismo como a las teorías de la mente existentes por no tener en cuenta los puntos de vista, incluidos los que admiten los *qualia* (irreductibles).

acaso llegamos o cuando lleguemos a saber cuál es esta propiedad, seguiremos sin saber cómo es *ser* Alfredo, pero sabremos más *acerca de* él. Ningún conocimiento acerca de Alfredo, sea físico o de otra índole, será conocimiento de Alfredo "desde el interior". No somos Alfredo. Por consiguiente, hay todo un conjunto de trozos de conocimiento que se expresan mediante formas verbales como "que soy *yo mismo* el que está..." que Alfredo tiene y que nosotros simplemente no podemos tener porque no somos él.[11]

Cuando Alfredo ve el color que sólo él puede ver, una cosa que sabe es la manera en que su experiencia de dicho color difiere de su experiencia de ver rojo, y demás, *otra* es que él mismo está viéndolo. Los fisicalistas y también los fanáticos de los *qualia* deberían reconocer que ninguna información de ninguna clase que los *demás* tengan *acerca de* Alfredo es conocimiento del segundo tipo. Mi queja, sin embargo, concernía al primer tipo y era que la cualidad especial de su experiencia es ciertamente un hecho acerca de ella, y es un hecho que el fisicalismo deja fuera porque ninguna cantidad de información física nos dice qué es.

Nagel habla como si el problema que está planteando fuera el de extrapolar del conocimiento de una experiencia a otra, de imaginar cómo sería una experiencia desconocida sobre la base de otras conocidas. En términos del ejemplo de Hume, del conocimiento de algunos tonos de azul podemos extraer cómo sería ver otros tonos de azul. Nagel argumenta que el problema con los murciélagos *et al.* es que son demasiado distintos de nosotros. Es difícil ver en esto una objeción al fisicalismo. El fisicalismo no hace ninguna afirmación especial acerca de las facultades imaginativas o de extrapolación de los seres humanos, y es difícil ver por qué tendría que hacerlo.[12]

De cualquier modo, nuestro argumento del conocimiento no hace ninguna suposición sobre este tema. Si el fisicalismo fuera cierto, suficiente información acerca de Alfredo obviaría

[11] Conocimiento *de se*, para usar los términos que emplea Lewis 1979.

[12] Véanse los comentarios de Laurence Nemirow sobre "Cómo es ser un murciélago" en su reseña de 1980. En este caso, estoy en deuda en particular con una discusión sostenida con David Lewis.

cualquier necesidad de extrapolar o de realizar actos especiales de imaginación o de entendimiento a fin de saber todo acerca de su especial experiencia del color. *La información ya estaría en nuestras manos*. Pero claramente no lo está. Ése era el meollo del argumento.

4. *El terror del epifenomenalismo*

¿Hay alguna razón realmente *buena* para rehusarse a aceptar la idea de que los *qualia* son causalmente impotentes con respecto al mundo físico? Defenderé la respuesta de que no la hay, pero al hacerlo no diré nada acerca de dos puntos de vista asociados con la posición epifenomenalista clásica. El primero es que los *estados* mentales son ineficaces con respecto al mundo físico. Todo lo que me propongo defender es que es posible sostener que ciertas *propiedades* de ciertos estados mentales, a saber, aquellas a las que he llamado *qualia*, son tales que su posesión o su ausencia no constituye ninguna diferencia para el mundo físico. El segundo es que lo mental es *totalmente* ineficaz causalmente. Pese a todo lo que voy a decir, puede ser que se tenga que sostener que la instanciación de los *qualia* afecta a *otros estados mentales* aunque no a algo físico. En efecto, las consideraciones generales que tienen que ver con la manera como se podría llegar a tomar conciencia de la instanciación de los *qualia* sugieren esa posición.[13]

Normalmente se dan tres razones para sostener que un *quale* como lo doloroso de un dolor debe ser causalmente eficaz en el mundo físico, y así, por ejemplo, que su instanciación a veces tiene que conllevar una diferencia en lo que ocurre en el cerebro. Sostendré que ninguna tiene una fuerza real. (Estoy muy agradecido con Alec Hyslop y John Lucas por haberme convencido de esto.)

(i) Se supone que es simplemente obvio que lo doloroso del dolor es en parte responsable de que el sujeto busque evitar el dolor, diga "Duele", y demás. Pero, para invertir lo que dice Hume, una cosa puede no causar nada. Sea cual sea la frecuencia con que *B* sigue a *A*, e independientemente de lo obvia que

[13] Véase mi reseña de Campbell 1970, en Jackson 1972.

parezca al principio la causalidad de la conexión, la hipótesis de que *A* causa *B* puede invalidarse con una teoría englobadora que muestre que ambos son efectos distintos de un proceso causal común subyacente.

Para los no instruidos, la imagen en la pantalla del puño de Lee Marvin moviéndose de izquierda a derecha seguida inmediatamente por la imagen de la cabeza de John Wayne moviéndose en la misma dirección general parece tan causal como cualquier otra.[14] Y, desde luego, en innumerables películas de vaqueros hay imágenes semejantes a la primera seguidas de imágenes semejantes a la segunda. Todo esto no cuenta precisamente mucho cuando conocemos la teoría englobadora concerniente a cómo las imágenes pertinentes son ambas efectos de un proceso causal subyacente que involucra al proyector y a la película. El epifenomenalista puede decir exactamente lo mismo acerca de la conexión, digamos, entre lo doloroso y el comportamiento. Es simplemente una consecuencia del hecho de que ciertos acontecimientos del cerebro causan ambos.

(ii) La segunda objeción se relaciona con la teoría de la evolución de Darwin. Según la selección natural, los rasgos que evolucionan con el paso del tiempo son los que favorecen la supervivencia física. Podemos suponer que los *qualia* evolucionaron con el tiempo —nosotros los tenemos, las primeras formas de vida no— y así deberíamos esperar que los *qualia* favorecieran la supervivencia. La objeción es que difícilmente nos ayudarían a sobrevivir si no tienen ningún efecto en el mundo físico.

El atractivo de este argumento es innegable, pero hay una buena réplica para el mismo. Los osos polares tienen pelajes particularmente espesos y calientes. La teoría de la evolución explica esto (suponemos) señalando que tener un pelaje espeso y caliente favorece la supervivencia en el Ártico. Pero tener un pelaje espeso se acompaña de tener un pelaje pesado, y tener un pelaje pesado *no* favorece la supervivencia. Hace que el animal sea más lento.

¿Significa esto que hemos refutado a Darwin porque hemos encontrado un rasgo evolucionado —tener un pelaje pesado—

[14] *Cfr.* Piaget 1977.

que no favorece la supervivencia? Desde luego que no. Tener un pelaje pesado es inevitablemente concomitante de tener un pelaje caliente (en el contexto, el aislamiento moderno no existía), y las ventajas de tener un pelaje caliente para sobrevivir compensan con creces las desventajas de tener un pelaje pesado. La cuestión es que todo lo que podemos extraer de la teoría de Darwin es que deberíamos esperar que cualquier característica evolucionada *o bien* favoreciera la supervivencia *o bien* fuera un producto secundario de alguna otra que la favoreciera. El epifenomenalista sostiene que los *qualia* caen dentro de esta última categoría. Son un producto secundario de ciertos procesos cerebrales que son muy propicios para la supervivencia.

(iii) La tercera objeción se basa en una cuestión acerca de cómo llegamos a saber de otras mentes. Sabemos acerca de otras mentes al conocer el comportamiento de esos otros, al menos en parte. La naturaleza de la inferencia es un asunto de cierta controversia, pero lo que no es asunto de controversia es que procede del comportamiento. Ésa es la razón por la que pensamos que las piedras no sienten y los perros sí. Pero, dice la objeción, cómo puede el comportamiento de una persona ofrecer alguna razón para creer que tiene *qualia* como los míos, o de hecho cualquier tipo de *qualia*, a menos que su comportamiento se vea como *consecuencia* de los *qualia*. Las huellas del hombre Viernes eran prueba del hombre Viernes porque las huellas son consecuencias causales de pies que pertenecen a personas. Y un epifenomenalista no puede considerar al comportamiento, ni a algo físico, como el resultado de los *qualia*.

Pero considérese que leo en *The Times* que el equipo Spurs de Tottenham ganó. Esto ofrece excelentes pruebas de que *The Telegraph* también ha informado que los Spurs ganaron, pese al hecho de que (confío en ello) *The Telegraph* no obtiene los resultados de *The Times*. Cada uno de estos diarios envía a sus reporteros al juego. La noticia que da *The Telegraph* no es en ningún sentido consecuencia de la que da *The Times*, pero en cualquier caso, esto último ofrece buenas pruebas de lo primero.

El razonamiento que esto involucra puede reconstruirse de la siguiente manera. Leo en *The Times* que los Spurs ganaron. Esto me da razón para pensar que los Spurs ganaron porque

sé que la victoria de los *Spurs* es el candidato más probable a ser lo que causó que apareciera la noticia en *The Times*. Pero también sé que la victoria de los Spurs habría tenido muchos efectos, incluyendo casi con toda certeza una noticia en *The Telegraph*.

Mi razonamiento va de un efecto a su causa y de ahí de nuevo hacia otro efecto. El hecho de que ninguno de los efectos cause el otro carece de pertinencia. Ahora bien, el epifenomenalista reconoce que los *qualia* son efectos de lo que ocurre en el cerebro. Los *qualia* no causan nada físico pero son causados por algo físico. Por lo tanto, el epifenomenalista puede razonar yendo del comportamiento de los otros a los *qualia* de esos otros mediante un razonamiento que vaya del comportamiento de los otros hacia sus causas en los cerebros de los otros y luego de nuevo hacia sus *qualia*.

Por una u otra razón, uno bien puede pensar que ésta es una cadena de razonamiento más discutible que su modelo del caso de las noticias de los diarios. Y tendría razón. El problema de otras mentes es uno de los problemas filosóficos de mayor importancia, en cambio el problema de las noticias de otros diarios no lo es. Pero aquí no hay ningún problema especial de epifenomenalismo en cuanto opuesto, digamos, al interaccionismo.

Hay una respuesta muy comprensible a las tres réplicas que acabo de hacer. "Está bien, no hay refutación demoledora de la existencia de los *qualia* epifenoménicos. Pero el hecho es que son una excrecencia. No *hacen* nada ni *explican* nada; sirven meramente para calmar las intuiciones de los dualistas, y sigue siendo un misterio total cómo encajan en la visión del mundo de la ciencia. En pocas palabras, ni entendemos ni podemos entender su cómo y su porqué."

Esto es absolutamente cierto; pero no es ninguna objeción a los *qualia*, pues se apoya en una visión demasiado optimista del animal humano y sus facultades. Somos producto de la evolución. Entendemos y sentimos lo que tenemos que entender y sentir para sobrevivir. Los *qualia* epifenoménicos carecen totalmente de pertinencia para la supervivencia. En ninguna etapa de nuestra evolución, la selección natural favoreció a quienes han logrado entender cómo son causados, y las leyes que los

gobiernan o, de hecho, por qué existen. Y esto explica por qué no podemos entenderlo.

No se ha comprendido lo suficiente que el fisicalismo es una visión extremadamente optimista de nuestras facultades. Si es verdadero, tenemos, en líneas muy generales desde luego, una comprensión de nuestro sitio en el esquema de las cosas. Ciertos asuntos de auténtica complejidad nos vencen –hay muchísimas neuronas–, pero en principio lo tenemos todo. Pero considérese la probabilidad antecedente de que todo lo que hay en el Universo sea de una clase de algún modo pertinente para la supervivencia del *homo sapiens*. Será muy baja sin duda. Pero entonces uno tiene que admitir que es muy probable que haya una parte del esquema completo de las cosas, tal vez una parte grande, que jamás ninguna evolución nos acercará a conocer o entender. Por la sencilla razón de que ese conocimiento y esa comprensión carecen de pertinencia para la supervivencia.

Los fisicalistas suelen hacer hincapié en que, según su opinión, somos una parte de la naturaleza, lo que es bastante justo. Pero si somos una parte de la naturaleza, somos como la naturaleza nos ha dejado ser después de quién sabe cuántos años de evolución, y cada peldaño de esa progresión evolutiva ha sido una cuestión de azar limitada sólo por la necesidad de preservar o aumentar el valor de supervivencia. Lo increíble es que entendamos todo lo que entendemos; y no es de sorprender que haya asuntos que escapan a nuestra comprensión. Quizá uno de ellos sea saber exactamente cómo encajan los *qualia* epifenoménicos en el esquema de las cosas.

Esto puede parecer una visión indebidamente pesimista de nuestra capacidad para articular una imagen verdaderamente global de nuestro mundo y del lugar que ocupamos en él. Pero supongamos que descubriéramos que en el fondo de los más profundos océanos vive una especie de babosa marina que diera signos de ser inteligente. Tal vez la supervivencia en esas condiciones requiriera facultades racionales. Pese a su inteligencia, estas babosas marinas tienen sólo una concepción muy limitada del mundo en comparación con la nuestra, lo que se explica por la naturaleza de su entorno inmediato. No obstante, han desarrollado ciencias que funcionan sorprendentemente bien en estos términos limitados. También tienen filósofos a los que

llaman babosistas. Algunos se autonombran babosistas de línea dura, y otros confiesan ser babosistas suaves.

Los babosistas de línea dura sostienen que los términos limitados (o algunos muy parecidos a ellos que puedan irse introduciendo conforme sus ciencias progresen) bastan en principio para describir todo sin dejar nada fuera. En ciertos momentos de debilidad, estos babosistas de línea dura reconocen que tienen la impresión de que su teoría está dejando algo fuera. Luchan contra esta idea y contra sus oponentes, los suaves, señalando —con toda razón— que nunca ningún babosista ha logrado descifrar cómo encaja este misterioso residuo en la muy exitosa visión que sus ciencias han desarrollado, y siguen desarrollando, de cómo funciona su mundo.

Nuestras babosas marinas no existen, pero podrían existir; y también podrían existir superseres que están en relación con nosotros como nosotros estamos en relación con las babosas marinas. No podemos adoptar la perspectiva de estos superseres, porque no somos ellos, pero la posibilidad de una perspectiva así es, en mi opinión, un antídoto contra el optimismo excesivo.

[*Traducción de Laura E. Manríquez*]

BIBLIOGRAFÍA

Campbell, K., 1976, *Metaphysics*, Dickenson, Belmont.

———, 1970, *Body and Mind*, Anchor Books, Nueva York. [Versión en castellano: *Cuerpo y mente*, trad. S. Marín, Instituto de Investigaciones Filosóficas-UNAM, México, 1987.]

Corniman, J.W., 1971, *Materialism and Sensations*, Yale University Press, New Haven y Londres.

Dennett, D.C., 1978, "Current Issues in the Philosophy of Mind", *American Philosophical Quarterly*, vol. 15, pp. 249–261.

Feyerabend, P. y G. Maxwell (comps.), 1966, *Mind, Matter, and Method*, University of Minnesota Press, Mincápolis.

Jackson, F., 1972, Reseña de *Body and Mind* de K. Campbell, *Australasian Journal of Philosophy*, vol. 50, no. 1, pp. 77–80.

Kirk, R., 1979, "From Physical Explicability to Full-Blooded Materialism", *The Philosophical Quarterly*, vol. 29, pp. 229–237.

Kirk, R., 1974, "Sentience and Behaviour", *Mind*, vol. 83, pp. 43–60.
Land, E.H., 1959, "Experiments in Color Vision", *Scientific American*, vol. 200, pp. 84–99.
Lewis, D., 1979, "Attitudes *De Dicto* and *De Se*", *Philosophical Review*, vol. 88, pp. 513–543.
Locke, D., 1976, "Zombies, Schizophrenics and Purely Physical Objects", *Mind*, vol. 85, no. 337, pp. 97–99.
Lycan, W.G., 1979, "A New Lilliputian Argument Against Machine Functionalism", *Philosophical Studies*, vol. 35, pp. 279–287.
Meehl, E.P., 1966, "The Compleat Autocerebroscopist", en Feyerabend y Maxwell 1966, pp. 103–180.
Mellor, D.H., 1973, "Materialism and Phenomenal Qualities", *Aristotelian Society*, vol. 47 sup., pp. 107–119.
Nagel, T., 1974, "What Is It Like to Be a Bat?", *Philosophical Review*, vol. 83, pp. 435–450. [Incluido en esta antología, pp. 45–63.]
Nemirow, L., 1980, Reseña de *Mortal Questions* de T. Nagel, *Philosophical Review*, vol. 89, pp. 473–477.
Piaget, J., 1977, "The Child's Conception of Physical Causality", reimpreso en E. Gruber y J. Vonèche (comps.), 1977, *The Essential Piaget*, Routledge and Kegan Paul, Londres.
Shoemaker, S., 1975, "Functionalism and Qualia", *Philosophical Studies*, vol. 27, pp. 291–315. [Incluido en esta antología, pp. 183–212.]
Wells, H.G., s.f., *The Country of the Blind and Other Stories*, s.e., Londres.

LO QUE MARÍA NO SABÍA*

FRANK JACKSON

María está confinada en una habitación blanca y negra, ha sido educada con libros en blanco y negro y mediante lecciones transmitidas en un televisor en blanco y negro. De esta manera aprende todo lo que hay que saber sobre la naturaleza física del mundo. Conoce todos los hechos físicos acerca de nosotros y nuestro entorno, en un sentido amplio de "físico" que incluye todo lo que hay según la física, la química y la neurofisiología *acabadas*, y todo lo que hay que saber acerca de los hechos causales y relacionales resultado de todo esto, incluyendo, por supuesto, los papeles funcionales. Si el fisicalismo es verdadero, María sabe todo lo que hay que saber; pues suponer lo contrario equivale a suponer que hay más por conocer aparte de todos los hechos físicos, y esto es justamente lo que el fisicalismo niega.

El fisicalismo no es la tesis no polémica de que el mundo actual es en gran medida físico, sino la tesis desafiante de que es totalmente físico. Ésta es la razón por la que los fisicalistas tienen que sostener que el conocimiento físico completo es conocimiento completo *simpliciter*. Pues supongamos que no es completo: entonces nuestro mundo debe diferir de otro mundo, $M(P)$, del cual es completo, y la diferencia tiene que radicar en hechos no físicos; pues nuestro mundo y $M(P)$ coinciden en

* Debo mucho a las discusiones sostenidas con David Lewis y con Robert Pargetter.

todas las cuestiones físicas. Por lo tanto, el fisicalismo sería falso en nuestro mundo (aunque de un modo contingente, pues sería verdadero en $M(P)$).[1]

Sin embargo, parece que María no sabe todo lo que hay que saber, pues cuando salga de la habitación en blanco y negro o cuando le den un televisor a color, aprenderá, por ejemplo, cómo es ver algo rojo. Esto se describe correctamente como *aprendizaje* —María no dirá "ah sí". Por lo tanto, el fisicalismo es falso. Éste es el argumento del conocimiento en contra del fisicalismo en una de sus manifestaciones.[2] Este texto es una réplica a tres objeciones presentadas por Paul M. Churchland en contra de dicho argumento.[3]

1. *Tres aclaraciones*

El argumento del conocimiento no se apoya en la dudosa afirmación de que lógicamente no es posible imaginar cómo es tener la sensación de rojo a menos que se haya tenido la sensación de rojo. Los poderes de la imaginación no tienen nada que ver aquí. La discusión en torno a María no es que, pese a sus fantásticos conocimientos de neurofisiología y todo lo físico, no *podría imaginar* cómo es tener la sensación de rojo; es que, de hecho, *no lo sabría*. Pero si el fisicalismo es cierto, lo sabría; y no haría falta recurrir a grandes poderes imaginativos. La imaginación es una facultad en la que tienen que apoyarse quienes *carecen* de conocimiento.

En segundo lugar, tampoco la intensionalidad del conocimiento tiene nada que ver en este caso. El argumento no se apoya en suponer falsamente que si S sabe que a es F y si $a = b$, entonces S sabe que b es F. Atañe más bien a la naturaleza de

[1] En este caso, la afirmación no consiste en que si el fisicalismo es verdadero, sólo lo que se expresa en un lenguaje explícitamente físico es conocimiento. Consiste en que si el fisicalismo es verdadero, entonces si alguien sabe todo lo expresado o expresable en un lenguaje explícitamente físico, ese alguien lo sabe todo. *Pace* Horgan 1984.

[2] Concretamente la de mi artículo "*Qualia* epifenoménicos" (Jackson 1982) [incluido en esta antología, pp. 95–110]. Véase también Nagel 1974 [incluido en esta antología, pp. 45–63]; y Robinson 1982.

[3] Churchland 1985. Las referencias de página que se citan en adelante son de este artículo, a menos que se señale algo distinto.

todo el cuerpo de conocimiento que María tiene antes de ser liberada: ¿está completo o se le escapan algunos hechos? Lo que sí tiene que ver es que *S* puede saber que *a* es *F* y *saber* que $a = b$, aunque posiblemente no sepa que *b* sea *F*, en virtud de no estar suficientemente alerta lógicamente para sacar las consecuencias. Si la falta de conocimiento por parte de María fuera siquiera de este tipo, no habría en ella ninguna amenaza para el fisicalismo. Pero es muy difícil creer que su falta de conocimiento podría remediarse simplemente con que de un modo explícito sacara suficientes consecuencias lógicas de su muy extenso conocimiento físico. Dotarla de una gran perspicacia lógica y de persistencia en sí mismo no basta para llenar las lagunas que hay en su conocimiento. Al ser liberada, no dirá: "Podría haber descubierto todo esto antes haciendo algunas inferencias puramente lógicas adicionales."

En tercer lugar, el conocimiento del que María carecía y que es de particular importancia para el argumento del conocimiento en contra del fisicalismo es *el conocimiento acerca de las experiencias de los demás*, no acerca de sus propias experiencias. Cuando es liberada, tiene nuevas experiencias, experiencias del color que nunca antes había tenido. Por consiguiente, no es una objeción al fisicalismo que ella aprenda *algo* al ser liberada. Antes de que la dejaran salir, no podría haber conocido hechos acerca de su experiencia de rojo, pues no había hechos de ese tipo que pudiera conocer. En esto pueden estar de acuerdo tanto los fisicalistas como los no fisicalistas. Después de que es liberada, las cosas cambian; y el fisicalismo puede fácilmente admitir que ella aprende esto; después de todo, algunas cosas físicas cambiarán, por ejemplo, sus estados cerebrales y sus papeles funcionales. El problema para el fisicalismo es que, después de que María vea su primer tomate maduro, se dará cuenta de lo empobrecida que había sido *durante todo ese tiempo* su concepción de la vida mental de los *otros*. Se percatará de que, mientras ella realizaba sus minuciosas investigaciones de la neurofisiología de los otros y de los papeles funcionales de sus estados internos, había algo acerca de esta gente de lo que no era en absoluto consciente. Durante todo ese tiempo, sus experiencias (o muchas de ellas, las obtenidas de los tomates, del cielo,...) tuvieron un rasgo conspicuo para ellos pero hasta entonces oculto para ella (de

hecho, y no por cuestión de lógica). Pero conocía todos los hechos físicos acerca de ellos todo el tiempo; por lo tanto, lo que no sabía hasta que fue liberada no es un hecho físico acerca de las experiencias de los otros. Pero sí es un hecho acerca de ellos. Ése es el problema para el fisicalismo.

2. *Las tres objeciones de Churchland*

(i) La primera objeción de Churchland es que el argumento del conocimiento contiene un defecto que "es la simplicidad en sí misma" (p. 23). El argumento es ambiguo sobre el sentido de "saber acerca de". ¿Cómo? Churchland propone que la siguiente es "una versión convenientemente ajustada" del argumento del conocimiento:

(1) María sabe todo lo que hay que saber acerca de los estados cerebrales y sus propiedades.

(2) No es el caso que María sepa todo lo que hay que saber acerca de las sensaciones y sus propiedades.

Por lo tanto, por la ley de Leibniz,

(3) Las sensaciones y sus propiedades $\neq$ los estados cerebrales y sus propiedades (p. 23).

Churchland señala, de un modo bastante plausible, que el tipo o la clase de conocimiento que implica la premisa 1 difiere de la clase de conocimiento que implica la premisa 2. Podríamos seguir su ejemplo y llamar al primero "conocimiento por descripción", y al segundo "conocimiento por familiaridad"; pero sean cuales sean los epítetos, tiene razón en que el argumento expuesto implica un uso muy dudoso de la ley de Leibniz.

Mi réplica es que el argumento expuesto puede ser conveniente, pero no es exacto. No es el argumento del conocimiento. Tomemos, por ejemplo, la premisa 1. La esencia del argumento del conocimiento radica en que María, antes de ser liberada, *no* sabe todo lo que hay que saber acerca de los estados cerebrales y sus propiedades, porque no sabe acerca de ciertos *qualia* asociados con ellos. Lo que es completo, según el argumento, es su conocimiento de las cuestiones físicas. Una manera conveniente y exacta de exponer el argumento es la siguiente:

(1)′ María (antes de ser liberada) sabe todo lo físico que hay que saber acerca de otras personas.

(2)′ María (antes de ser liberada) no sabe todo lo que hay que saber acerca de otras personas (porque ella aprende algo acerca de ellos al ser liberada).

Por consiguiente,

(3)′ Existen verdades acerca de los demás (y de ella misma) que escapan a la explicación fisicalista.

Lo que directamente viene al caso no es la clase ni la manera ni el tipo de conocimiento que tiene María, sino *lo que* sabe. Lo que sabe de antemano es *ex* hipótesis todo lo físico que hay que saber pero, ¿es todo lo que hay que saber? Ésa es la pregunta crucial.

Hay, no obstante, un desafío pertinente que involucra cuestiones acerca de las clases de conocimiento. Concierne *al apoyo* a la premisa 2′. El caso en favor de la premisa 2′ es que María aprende algo al ser liberada, adquiere conocimiento y esto implica que su conocimiento anterior (*lo que* sabía, independientemente de que fuera por descripción, por familiaridad, o como sea) era incompleto. El reto, presentado por David Lewis y Laurence Nemirow, es que, al ser liberada, María no aprende nada ni adquiere conocimiento en el sentido pertinente. Lo que María adquiere al ser liberada es cierta habilidad representacional o imaginativa; es un conocer cómo más que un conocer que. Por consiguiente, un fisicalista puede aceptar que María adquiere algo muy significativo que cae dentro del rubro del conocimiento —lo que difícilmente se podría negar—, sin admitir que esto muestra que su conocimiento fáctico previo era imperfecto. Sabía todo *lo que* de antemano había que saber acerca de las experiencias de los demás, pero carecía de una habilidad hasta ser liberada.[4]

[4] Véanse Nemirow 1980; y Lewis 1983. Churchland menciona a Nemirow y a Lewis, y puede ser que tuviera la intención de que su objeción fuese esencialmente la que acabo de ofrecer. Sin embargo, de manera bastante explícita dice (al final de la p. 23) que su objeción no necesita un análisis del conocimiento pertinente basado en "habilidades".

Ahora bien, es ciertamente verdadero que María adquirirá habilidades de varias clases después de su liberación. Por ejemplo, podrá imaginarse cómo es ver rojo, podrá recordar cómo es verlo y podrá entender por qué sus amigos la consideraban tan en desventaja (algo que, antes de ser liberada, siempre la había desconcertado). Pero, ¿es plausible decir que eso es *todo* lo que adquirirá? Supóngase que recibió una lección sobre el escepticismo acerca de otras mentes mientras estuvo prisionera. Al ser liberada, ve un tomate maduro en condiciones normales, y así tiene una sensación de rojo. Su primera reacción es decir que ahora sabe más acerca de la clase de experiencias que tienen los otros cuando observan tomates maduros. Luego recuerda la lección y se empieza a preocupar. ¿Sabe realmente más acerca de cómo son las experiencias de los demás, o se está permitiendo hacer una generalización totalmente fortuita a partir de un caso? Al final decide que sí sabe y que el escepticismo está equivocado (aun cuando, como muchos de nosotros, no esté segura de cómo demostrar sus errores). ¿Por qué daba tantas vueltas?, ¿por sus habilidades? Desde luego que no; sus habilidades representacionales fueron una constante conocida en todo momento. ¿Acerca de qué otra cosa entonces se estuvo rompiendo la cabeza sino acerca de si había obtenido o no conocimiento fáctico acerca de los demás? No habría nada sobre lo cual romperse la cabeza si *todo* lo que hubiese adquirido al ser liberada fuera una habilidad.

Reconozco que no tengo ninguna *prueba* de que, al ser liberada, María adquiera, además de habilidades, conocimiento fáctico acerca de las experiencias de los demás —y no sólo porque no tengo refutación del escepticismo. Mi afirmación es que el argumento del conocimiento es un argumento válido a partir de premisas altamente plausibles —aunque, hay que admitirlo, no demostrables—, premisas para la conclusión de que el fisicalismo es falso. Y eso, después de todo, es una objeción todo lo buena que uno podría esperar en esta área de la filosofía.

(ii) La segunda objeción de Churchland (pp. 24–25) es que tiene que haber algo equivocado en el argumento, pues prueba demasiado. Supóngase que María recibió una serie especial de lecciones por medio de su televisor en blanco y negro de un

dualista auténtico, que le explicaban las "leyes" que gobiernan el comportamiento del "ectoplasma" y le hablaban acerca de los *qualia*. Esto no afectaría la verosimilitud de la afirmación de que al ser liberada aprende algo. Así, si el argumento funciona en contra del fisicalismo, también funciona en contra del dualismo.

Mi réplica es que las lecciones acerca de los *qualia* a través del televisor en blanco y negro no le dicen a María todo lo que hay que saber acerca de los *qualia*. Quizás le digan algunas cosas acerca de los *qualia*, por ejemplo, que no aparecen en la explicación fisicalista, y que el *quale* para el que usamos el "amarillo" es aproximadamente tan distinto de aquel para el que usamos "azul", como lo es el negro del blanco. Pero, ¿por qué habría que suponer que le dicen todo acerca de los *qualia*? Por otro lado, resulta plausible que las clases por medio del televisor en blanco y negro puedan en principio revelarle a María todo lo que hay en la explicación fisicalista. No necesitamos un televisor a color para aprender física o psicología funcionalista. Para obtener un buen argumento en contra del dualismo (el dualismo de atributos; el caso del ectoplasma es sólo un juego), la premisa del argumento del conocimiento acerca de que María sabe todo según el fisicalismo antes de su liberación tiene que ser reemplazada por la premisa de que sabe todo según el dualismo. La primera es plausible; la segunda no. Por lo tanto, no hay problema de "paridad de razones" para los dualistas que usan el argumento del conocimiento.

(iii) La tercera objeción de Churchland es que el argumento del conocimiento sostiene "que María ni siquiera podría *imaginarse* cómo serían las experiencias pertinentes, pese a su exhaustivo conocimiento neurocientífico, y, por lo tanto, le debe seguir faltando cierta información crucial" (p. 25), una afirmación contra la cual Churchland argumenta en seguida.

Pero, como lo subrayamos antes, el argumento del conocimiento afirma que María no sabría cómo es la experiencia pertinente. Lo que podría imaginar es otro asunto. Si su conocimiento es imperfecto, a pesar de que contenga todo lo que

hay que saber según el fisicalismo, entonces el fisicalismo es falso, sean cuales sean sus poderes imaginativos.

[*Traducción de Laura E. Manríquez*]

BIBLIOGRAFÍA

Horgan, T., 1984, "Jackson on Physical Information and Qualia", *Philosophical Quarterly*, vol. 34, no. 135, pp. 147–152.

Churchland, P.M., 1985, "Reduction, Qualia, and the Direct Introspection of Brain States", *Journal of Philosophy*, vol. 82, no. 1, pp. 8–28.

Jackson, F., 1982, "Epiphenomenal Qualia", *Philosophical Quarterly*, vol. 32, no. 127, pp. 127–136. [Incluido en esta antología, pp. 95–110.]

Lewis, D., 1983, "Postscript to 'Mad Pain and Martian Pain' ", *Philosophical Papers*, vol. 1, Oxford University Press, Nueva York, pp. 130–132.

Nagel, T., 1974, "What Is It Like to Be a Bat?", *Philosophical Review*, vol. 83, no. 4, pp. 435–450. [Incluido en esta antología, pp. 45–63.]

Nemirow, L., 1980, Reseña de *Mortal Questions*, de T. Nagel, *The Philosophical Review*, vol. 89, no. 3, pp. 473–477.

Robinson, H., 1982, *Matter and Sense*, Cambridge University Press, Nueva York.

LO QUE ENSEÑA LA EXPERIENCIA*

DAVID LEWIS

1. *La experiencia es la mejor maestra*

Se dice que la experiencia es la mejor maestra, y que el aula no sustituye a la vida real. Hay algo de verdad en ello. Si uno quiere saber cómo es una experiencia nueva y diferente, la aprenderá saliendo a la calle y *teniendo* realmente esa experiencia. No puedes aprenderla por lo que te digan los demás acerca de ella, por detalladas que sean las lecciones que te den.

¿Prueba esto algo sobre la metafísica de la mente y los límites de la ciencia? Creo que no.

*Ejemplo: Zorrillos y Vegemite.*** He olido a los zorrillos, de modo que sé cómo es oler zorrillos. Pero los zorrillos sólo viven en algunas regiones del mundo, así que quizá nunca los hayas olido. Si no has olido a un zorrillo, entonces no sabes cómo huele. Y nunca lo sabrás, a menos que un día huelas personalmente a un zorrillo. Por otro lado, tal vez tú has probado la Vegemite, esa famosa sustancia australiana, y yo nunca lo he

* Parte de este escrito deriva de una conferencia dictada en la La Trobe University en 1981. Agradezco a La Trobe por su apoyo en 1981, a la Harvard University por su ayuda con la Beca Santayana en 1988, y a Frank Jackson por las utilísimas discusiones que hemos sostenido.

** Vegemite y Marmite (que se mencionan más adelante en el texto) son marcas comerciales de dos populares sazonadores vegetales fabricados a base de levadura. Vegemite es una marca australiana, mientras que Marmite es originaria del Reino Unido. [N. de las comps.]

hecho. De modo que tú sabes cómo sabe la Vegemite. Yo no, y a menos que la pruebe (¿y arruinar un buen ejemplo?), nunca lo sabré. No servirá de nada tomar lecciones sobre la composición química del olor del zorrillo o de la Vegemite, la fisiología de la nariz o de las papilas gustativas y la neurofisiología de los nervios sensoriales y del cerebro.

Ejemplo: La científica cautiva.[1] María, una brillante científica, ha vivido desde su nacimiento en una habitación donde todo es blanco y negro. (Ella misma se encuentra pintada de pies a cabeza.) Ve el mundo por una televisión en blanco y negro. A través de la televisión, lee libros, participa en debates, observa los resultados de los experimentos realizados bajo su dirección. De esta manera se vuelve la mayor experta del mundo sobre el color, la visión de los colores y los estados cerebrales producidos por estar expuesto a los colores. Pero ella no sabe cómo es ver colores. Y nunca lo sabrá, a menos que escape de su habitación.

Ejemplo: El murciélago.[2] El murciélago es una criatura extraña, provista de un sonar que es completamente distinto de cualquiera de nuestros sentidos. Nunca podremos tener las experiencias de un murciélago, porque no podríamos parecernos lo suficientemente a él para tener esas experiencias y seguir siendo nosotros mismos. Jamás sabremos cómo es ser un murciélago. Ni siquiera si llegáramos a conocer todos los hechos existentes acerca del comportamiento y las disposiciones conductuales del murciélago, sobre su estructura y sus procesos físicos, sobre su organización funcional. Ni siquiera si llegáramos a conocer todos los mismos tipos de hechos físicos acerca de todos los demás murciélagos, de otras criaturas o de nosotros. Ni siquiera si llegáramos a poseer todos los hechos físicos del mundo. Ni siquiera si fuéramos capaces de identificar todas las implicaciones matemáticas y lógicas de estos hechos, por complejas que sean y por lejos que se encuentren de la deducción finita.

La experiencia es la mejor maestra en este sentido: tener una experiencia constituye la mejor manera, o quizá la única mane-

[1] Véanse Jackson 1982, [incluido en esta antología, pp. 95–110]; y Jackson 1986 "Lo que María no sabía" [incluido en esta antología, pp. 111–118].

[2] Véase Farrell 1950, y Nagel 1974 "Cómo es ser un murciélago" [incluido en esta antología, pp. 45–63].

ra, de llegar a saber cómo es esa experiencia. Ningún acopio de información científica sobre los estímulos que la producen y el proceso que ocurre en ti mismo cuando la tienes te permitirá saber cómo es tener esa experiencia.

2. . . . *Pero no necesariamente*

Tener una experiencia es sin duda una buena forma, y seguramente la única forma práctica, de llegar a saber cómo es esa experiencia. ¿Podemos decir lisa y llanamente que es la única forma *posible*? Probablemente no. Se produce un cambio en ti cuando tienes la experiencia y llegas así a saber cómo es. Tal vez, en principio, se podría producir exactamente el mismo cambio por medio de una neurocirugía precisa mucho más allá de los límites de la técnica actual. O podría producirse por medio de la magia. Si hacemos caso omiso de las leyes de la naturaleza, que después de todo son contingentes, entonces no hay ninguna relación necesaria entre causa y efecto: cualquier cosa podría causar cualquier otra. Por ejemplo, un encantamiento podría producirle a uno exactamente el mismo efecto que le produciría el olor del zorrillo cuando se huele por primera vez. Podríamos polemizar acerca de si un estado producido de esta manera artificial merece el *nombre* de "saber cómo es oler un zorrillo", pero podemos imaginar que, en lo concerniente a lo que sucede dentro de uno, no sería en absoluto diferente.[3]

Así como podemos imaginar que un encantamiento es capaz de producir el mismo cambio que un olor, podemos igualmente imaginar que las lecciones de ciencia podrían ocasionar ese mismo cambio. Incluso esto es posible, en el sentido más amplio del término. Si ignoráramos todo lo que sabemos sobre la manera en que el mundo realmente funciona, no podríamos tener ninguna idea de lo que le sucedería a alguien a quien se le enseñara la química del olor y la fisiología de la nariz. Podría haber un mecanismo causal que transformara las lecciones de ciencia en lo que aquella experiencia nos proporciona, sea lo que esto fuere. Pero no lo hay. Que la experiencia es la mejor maestra para enseñarnos cómo es una nueva experiencia no

[3] Véase Unger 1966.

constituye una verdad absolutamente necesaria, es una verdad contingente. Pero tenemos buenos motivos para pensar que es verdad.

En cualquier caso, tenemos buenas razones para pensar que algo como lo anterior es verdad, pero menos razones para estar seguros de qué exactamente es verdad. Quizá cierto método de dar clases que todavía no ha sido inventado y cierta manera de tomarlas que aún no se ha puesto en práctica podrían darnos una gran sorpresa. Consideremos leer a primera vista: un músico preparado puede leer una partitura y saber cómo sería escuchar la música. De no estar yo enterado de que algunas personas son capaces de leer a primera vista, jamás lo hubiera creído humanamente posible. Desde luego, la moraleja es que la música nueva no es enteramente nueva; la gran experiencia nueva es un reordenamiento de una multitud de antiguas experiencias pequeñas. Podría ocurrir exactamente lo mismo con los nuevos olores y sabores en comparación con los antiguos, o incluso con la visión a colores en comparación con la visión en blanco y negro;[4] o incluso con la experiencia sensorial del sonar en comparación con el tipo de experiencia sensorial del que gozamos. Lo que sí podemos decir con cierta seguridad es que no tenemos la facultad para saber, con base en unas simples lecciones de ciencia, cómo sería una experiencia *suficientemente nueva*. Pero, ¿qué tan nueva es "suficientemente nueva"? A este respecto, podrían esperarnos algunas sorpresas.

3. *Tres maneras de no captar la idea*

La primera manera. Un literalista podría considerar la frase "saber cómo es" e interpretarla como "saber a qué se parece". Entonces podría preguntar: ¿Dónde está la dificultad? ¿Por qué no se te puede decir cuáles son las experiencias que se parecen entre sí? No es preciso haber tenido las experiencias; todo lo que necesitas, para que te enseñen las lecciones, es alguna forma de referirte a ellas. Te podrían decir: el olor de un zorrillo de algún modo se parece al olor de hule quemándose. Me han dicho que el sabor de la Vegemite se parece un poco al de la

[4] Para esta especulación, véase Churchland 1985.

Marmite. María, en su mundo negro y blanco, podría saber más que la mayoría de nosotros acerca de las semejanzas entre las experiencias del color. Podría saber cuáles son las que reciben espontáneamente el calificativo de "similares" por parte de los sujetos que las tienen, cuáles son los cambios graduales entre ellas que tienden a pasar inadvertidos, cuáles se confunden con qué otras en la memoria, cuáles entrañan la activación de, aproximadamente, las mismas neuronas a ritmos similares, y así sucesivamente. Podríamos incluso saber a qué se parecen las experiencias del sonar del murciélago con sólo saber que no se parecen en absoluto a ninguna experiencia de los seres humanos, pero que sí se parecen —quizá— a ciertas experiencias que se producen en algunos peces. Esto es no captar la idea. Con perdón del literalista, "saber cómo es" no significa "saber a qué se parece". A lo sumo es verdad que saber a qué se parece *puede* ayudar a saber cómo es. Si te enseñan que la experiencia *A* se parece mucho a *B* y *C*, menos a *D*, y absolutamente nada a *E*, esto te ayudará a saber cómo es *A*, si ya sabes cómo son *B*, *C*, *D* y *E*. De lo contrario, no te servirá de nada. No sabré ni un ápice más acerca de cómo es el sabor de la Vegemite si me dicen que sabe igual que la Marmite, porque tampoco sé cómo sabe la Marmite. (Y no sabré ni un ápice más acerca de cómo es el sabor de la Marmite si me dicen que sabe como la Vegemite.) Tal vez María sabe lo suficiente para triangular exactamente cada experiencia de color en una red de semejanzas, o en múltiples redes de semejanza según los diferentes aspectos, a pesar de que no sabrá nunca cómo es un nodo cualquiera de alguna de las redes. Tal vez podríamos hacer esto mismo con las experiencias de murciélago. Pero ningún acopio de información sobre las semejanzas nos ayudará, por sí solo, a saber cómo es una experiencia.

La segunda manera. Si no sé cómo sería conducir a toda velocidad una locomotora a vapor en una noche fría y tormentosa, parte de mi problema radica simplemente en que ignoro qué experiencias tendría. La caldera emite muchísimo calor, especialmente cuando el fogonero abre la portezuela para echar más carbón; por otro lado, la cabina tiene corrientes de aire y ofrece escasa protección contra el temporal. ¿Tendría yo mucho calor o mucho frío? ¿O ambas cosas, alternadamente? ¿O se me

congelaría el rostro y me arderían las piernas? Si supiera las respuestas a estas preguntas, sabría mucho mejor cómo sería conducir la locomotora. De modo que, quizá, "saber cómo es" sólo significa "saber qué experiencias tiene uno". Y nuevamente: ¿cuál es entonces el problema? ¿Por qué no basta que te digan las experiencias que tendrías si, por ejemplo, probaras la Vegemite? Tampoco en este caso es preciso haber tenido las experiencias; todo lo que necesitas, que las lecciones te enseñen, es una forma de referirte a ellas. Tenemos maneras de referirnos a experiencias que no hemos tenido. Podemos referirnos a ellas en términos de sus causas: la experiencia que uno tiene al probar la Vegemite, la experiencia que uno tiene al probar una sustancia de tal y cual composición química. O podemos referirnos a ellas en términos de sus efectos: la experiencia que causó que Alfredo exclamara: "¡Uuuuy!" O podemos referirnos a ellas en términos de los estados físicos del sistema nervioso que median entre esas causas y esos efectos: la experiencia que uno tiene cuando sus nervios son activados según tal y cual patrón. (Según algunos materialistas, y me cuento entre ellos, esto significa la experiencia que es idéntica a tal y cual patrón de activación. De acuerdo con otros materialistas, significa la experiencia que se realiza en tal y cual patrón de activación. De acuerdo con muchos dualistas, significa la experiencia que no es sino la compañera legítima de tal y cual patrón de activación. Pero cualquiera que sea, obtenemos una forma de referirnos a la experiencia.) María, en su mundo en blanco y negro, se encuentra en posición de referirse a experiencias del color en todas estas formas. Por lo tanto, no deberías tener ningún problema para comunicarle cuáles son exactamente las experiencias que tienes al ver los colores. O más bien, tu único problema consiste en que le estarías comunicando lo que ella ya sabe perfectamente. En general, saber qué es la *X* consiste en saber que la *X* es la *Y*, donde no resulta tan evidente que la *X* sea la *Y*. (Saber solamente que la *X* es la *X* no servirá, desde luego, porque resulta demasiado obvio.) Si María sabe que la experiencia de ver verde es la experiencia asociada con tal y cual patrón de activación nerviosa, conoce entonces el tipo correcto de identidad no evidente. De suerte que sabe qué experiencia tiene uno al ver verde.

(En ocasiones se sugiere que se necesita un "designador rígido": uno sabe qué es la *X* cuando sabe que la *X* es la *Y* sólo si "la *Y*" es un término cuyo referente no depende de ninguna situación contingente. En primer lugar, esta sugerencia es falsa. Se puede saber quién es el hombre que está en el balcón sabiendo que el hombre del balcón es el primer ministro, aun cuando ni "el primer ministro" ni ninguna otra frase disponible designe rígidamente al hombre que, de hecho, es el primer ministro. En segundo lugar, de acuerdo con una de las versiones del materialismo –la que yo acepto–, una descripción del tipo "el estado de tener una activación nerviosa con tal y cual patrón" *es* un designador rígido, y lo que designa es, de hecho, una experiencia; y de acuerdo con otra versión del materialismo, una descripción del tipo "estar en tal o cual estado que desempeña tal y cual papel funcional" es un designador rígido de una experiencia. Así, aun cuando se diera por sentada esa falsa sugerencia, no se demostraría, sin caer en una petición de principio contra el materialismo, que María no puede saber cuál es la experiencia que uno tiene al ver rojo.)

Como María *sí* sabe las experiencias que tendría si viera los colores, pero *no* sabe cómo sería ver los colores, será mejor concluir que "saber cómo es" no significa, después de todo, "saber qué experiencias tiene uno". El ejemplo de la locomotora resultó engañoso. Efectivamente, conociendo cuáles serían las experiencias que tendría el maquinista, yo puedo saber cómo sería conducir la locomotora, pero sólo porque ya sé cómo son esas experiencias. (Es importante que yo sepa cómo son bajo las descripciones adecuadas, como podría ser la descripción "rostro congelado y piernas que arden". Volveremos a esto más adelante.) María quizá sabe tan bien como yo que cuando el maquinista se asoma, exponiéndose a la tormenta, para ver las señales, tendrá la experiencia de ver a veces luces verdes y a veces luces rojas. Sabe mejor que yo cuáles experiencias tiene el maquinista cuando alcanza a ver las señales. Puede ofrecer muchas más descripciones equivalentes no obvias de esas experiencias de las que yo puedo ofrecer. Pero saber cuáles son las experiencias de color que tiene el maquinista no ayudará a María a saber cómo es su trabajo. Me ayudará a mí.

La tercera manera. Hasta que no vea verde, he aquí algo que María nunca sabrá: nunca sabrá que está viendo verde. La razón de ello es que hasta que no vea verde, nunca será verdad que está viendo verde. Hay cierto conocimiento que es irreductiblemente egocéntrico, o *de se*.[5] No es sólo el conocimiento de lo que sucede en el mundo; es el conocimiento de quién y cuándo es uno en el mundo. El conocimiento de lo que sucede en el mundo será igualmente verdadero para todos los que viven en ese mundo; mientras que el conocimiento egocéntrico puede ser verdadero para uno y falso para otro, o verdadero para uno en cierto momento y falso para esa misma persona en otro momento. Quizá María sabe desde antes, cuando planea su fuga, que a las nueve de la mañana del 13 de mayo de 1997 será el momento en que una persona previamente confinada en una celda en blanco y negro verá el color por primera vez. Pero hasta que ese momento llegue, nunca sabrá que ella misma está viendo el color, porque no lo está viendo. Lo que no es verdadero no es conocimiento. Y esto se aplica tanto al conocimiento egocéntrico como a cualquier otro conocimiento. De modo que sólo las personas de quienes es verdad una proposición egocéntrica pueden conocerla, y sólo pueden hacerlo cuando esa proposición es verdadera respecto de ellos. Que uno está viendo el color es una proposición egocéntrica. Por ende, hemos encontrado una proposición que María no puede conocer hasta que vea el color, que, ¡como de hecho ocurre, será el momento mismo en que sabrá por primera vez cómo es ver el color! ¿Hemos descubierto la razón por la cual la experiencia es la mejor maestra? ¿Y, después de todo, no en forma contingente, sino como una consecuencia necesaria de la lógica del conocimiento egocéntrico?

No; tenemos aquí dos fenómenos distintos, y sólo algún mágico conjuro sobre la "perspectiva de la primera persona" podría hacer que no notáramos la diferencia. En primer lugar, María seguirá probablemente sabiendo cómo es ver verde después de que deje de conocer la proposición egocéntrica de que está viendo verde en ese momento. Como lo que no es verdadero no es conocido, ella debe dejar de conocer esa proposición en cuanto deje de ver verde. (¿Quiere esto decir únicamente que

[5] Véase Lewis 1979, también incluido en Lewis 1983a, y Chisholm 1981.

deberíamos haber tomado una proposición egocéntrica diferente: la de que uno *ha visto* verde? No, porque en ese caso María podría seguir conociendo la proposición incluso después de haber olvidado cómo es ver verde, cosa que podría suceder si, al poco tiempo, la volvieran a encerrar.) En segundo lugar, María podría llegar a saber cómo es ver verde aunque no conociera la proposición egocéntrica. Podría no haber sabido anticipadamente que su ruta de escape iba a llevarla a través de un prado verde, y le podría tomar algo de tiempo reconocer la hierba por su forma. Así, al principio, quizá sólo sabría que estaba viendo uno u otro color y, por ello, descubriendo cómo eran las experiencias de uno u otro color, sin poder darle nombre ni a los colores ni a las experiencias. Sabría entonces cómo es ver verde, aunque no bajo esa descripción y, de hecho, bajo ninguna descripción que resultara más útil que "la experiencia del color que estoy teniendo ahora"; pero no conocería la proposición egocéntrica de que está viendo verde, ya que no sabría qué color está viendo. En tercer lugar, la obtención de un conocimiento egocéntrico puede tener prerrequisitos que no tienen nada que ver con la experiencia. Así como María no puede saber que está viendo verde hasta que efectivamente *ve* verde, no puede saber que está cumpliendo cincuenta años hasta que efectivamente *cumpla* cincuenta años. Pero –¡espero!– cumplir cincuenta años no entraña una experiencia especial. En suma, aunque de hecho uno puede obtener el conocimiento egocéntrico de que se halla en cierta situación sólo cuando se halla en ella, esto no es lo mismo que descubrir cómo es una experiencia sólo cuando uno tiene esa experiencia.

Acabamos de rechazar dos sugerencias que no funcionan de manera independiente, y podríamos observar que tampoco funcionan mejor cuando se las une. Uno sabe qué es la X cuando sabe que la X es la Y, donde la identidad no es demasiado obvia; y "la Y" podría ser una descripción egocéntrica. Por lo tanto, el conocimiento de que la X es la Y podría ser un conocimiento irreductiblemente egocéntrico, y en consecuencia un conocimiento que no puede tenerse hasta que sea verdadero de uno mismo que la X es la Y. De este modo, una forma de conocer qué es la X permanecerá fuera de tu alcance hasta que se vuelva verdadero de ti que la X es la Y. Una manera en que yo podría

obtener una identidad no obvia con respecto al sabor de la Vegemite sería que se volviera verdadero que el sabor de la Vegemite es el que yo estoy experimentando en este momento, ¡y que se volviera verdadero en el preciso momento en que yo pruebo Vegemite y descubro cómo es! ¿Se debe a esto que la experiencia sea la mejor maestra? No; los casos en los que se obtiene una identidad egocéntrica no obvia se encuentran por docenas, y la mayoría no tiene por resultado el descubrimiento de cómo es una experiencia. Supongamos que planeo por anticipado que ya no resistiré más y que probaré la Vegemite el próximo jueves al mediodía. Entonces, el mediodía del miércoles, si miro el reloj, obtendré el conocimiento egocéntrico poco obvio de que el sabor de la Vegemite es el que obtendré exactamente en veinticuatro horas, y así tengo una nueva forma de saber cómo es el sabor de la Vegemite. Pero el miércoles al mediodía todavía no sé cómo es. Otro ejemplo: de vez en cuando me encuentro junto a un catador de Vegemite. En esas ocasiones, y sólo en ellas, sé cómo es el sabor de Vegemite mediante el conocimiento de que es el sabor que está paladeando la persona que se encuentra a mi lado. Pero en ninguna de tales ocasiones ha ocurrido aún que yo sepa cómo es saborear Vegemite.

4. *La hipótesis de la información fenoménica*

No hay información física que María pueda recabar, en su mundo blanco y negro, que la ayude a saber cómo es ver los colores; no hay información física que podamos recabar sobre los murciélagos que nos ayude a saber cómo es tener sus experiencias; y lo mismo ocurre en otros casos. Hay una explicación natural y tentadora de por qué la información física no nos ayuda. Es la hipótesis de que, además de la información física, existe un tipo de información irreductiblemente distinta que se debe adquirir: la *información fenoménica*. Las dos son independientes. Dos casos posibles podrían ser exactamente iguales desde el punto de vista físico, pero diferir en el nivel fenoménico. Cuando adquirimos información física reducimos las posibilidades físicas, y quizá las reducimos hasta convertirlas en una sola, pero dejamos abierta una serie de posibilidades fenoménicas. Por otro lado, cuando tenemos una experiencia adquirimos información

fenoménica; se eliminan posibilidades que antes estaban abiertas, y en esto consiste aprender cómo es la experiencia.

(Analogía. Tomemos el problema relativo a la ubicación de un punto dentro de cierta zona del plano x–y. Se nos podría indicar que su coordenada x se encuentra dentro de determinados intervalos y fuera de otros. Podríamos incluso obtener suficiente información de este tipo para fijar precisamente la coordenada x. Pero ninguna información sobre x nos indicaría nada sobre la coordenada y; cualquier información sobre x sigue dejando abiertas todas las posibilidades de y. Pero cuando finalmente tomamos una medida de y, adquirimos una nueva clase de información; se eliminan posibilidades que antes estaban abiertas, y así es como sabemos dónde se encuentra el punto en el eje y.)

¿En que consiste la información fenoménica? Si la hipótesis de la información fenoménica es verdadera, entonces se tiene una respuesta fácil: se trata de información sobre la experiencia. En términos más específicos, se trata de información sobre cierta parte, aspecto o rasgo de la experiencia. Sin embargo, si la hipótesis es falsa, entonces sigue habiendo experiencia (con todas sus partes, aspectos y rasgos), pero ninguna información sobre la experiencia es información fenoménica. De modo que no puede decirse en términos neutrales, sin presuponer la hipótesis, que la información sobre la experiencia es información fenoménica. Pues si la hipótesis es falsa y el materialismo es verdadero, puede darse el caso de que toda la información existente sobre la experiencia sea información física, y bien puede estar contenida en las lecciones a las personas que no la han experimentado.

No cambia nada si se sustituye "experiencia" por alguna nueva frase de moda. Si en vez de "experiencia" dices "sensación cruda" (o simplemente "sensación"), o "la manera en que se siente", o "cómo es", sostengo que no estás significando nada distinto. ¿Hay algo que sea cómo es ser este robot? ¿Tiene este robot experiencias? No distingo ninguna diferencia entre la nueva pregunta y la vieja. ¿Sientes una quemadura de sol en la misma forma en que la siento yo? ¿Tenemos la misma sensación cruda? ¿Tenemos la misma experiencia cuando nos quema

el sol? También en este caso es la misma pregunta. "Conocer la sensación" y "saber cómo es" son intercambiables. (Salvo por el hecho de que lo primero podría insinuar una alternativa a la hipótesis de la información fenoménica.) De modo que si un partidario de la información fenoménica dice que su tema son las sensaciones crudas, o las maneras de sentir o cómo es, le responderé tal como lo haría si dijera que el tema es la experiencia. Quizá lo sea, *si* la hipótesis de la información fenoménica es verdadera; pero si ésta es falsa y el materialismo verdadero, aún sigue habiendo información sobre las sensaciones crudas, las maneras de sentir o el cómo es; pero en tal caso se trata de información física y puede ser transmitida en unas lecciones.

Podríamos obtener un posible candidato para la información fenoménica que no sea sólo la experiencia con un nuevo nombre; pero seguiría siendo tendencioso. Por ejemplo, se nos podría decir que la información fenoménica trata del carácter intrínseco de la experiencia. Un partidario de la información fenoménica podría creer realmente que ésta revela ciertas propiedades intrínsecas, no físicas, especiales de la experiencia. Podría incluso creer que revela la existencia de una cosa o un proceso especial, no físico, cuyas propiedades intrínsecas son *todas* no físicas. Pero no es, en modo alguno, el único en decir que la experiencia posee un carácter intrínseco. Muchos de los materialistas también lo decimos. Afirmamos que cierta experiencia de color es un estado, cualquiera que éste sea, que desempeña determinado papel funcional. De suerte que si lo que desempeña ese papel (universalmente, o para el caso de los seres humanos, o para el caso de algunos seres humanos) es cierto patrón de activación neuronal, el patrón de activación *es* entonces la experiencia (en el caso en cuestión). Por lo tanto, el carácter intrínseco de la experiencia es el carácter intrínseco del patrón de activación. Por ejemplo, una frecuencia de activación forma parte del carácter intrínseco de la experiencia. Si los materialistas estamos en lo correcto acerca de qué es la experiencia, entonces María, en su mundo en blanco y negro, sabe todo lo referente al carácter intrínseco de la experiencia del color, mientras que la mayoría de las personas que saben cómo es

la experiencia del color siguen sumidas en la total ignorancia sobre su carácter intrínseco.[6]

Decir que la información fenoménica concierne a los "*qualia*" sería casi igualmente tendencioso. Pues, ¿cómo se introdujo esta noción? A menudo, de la siguiente manera. Se nos pide que imaginemos a una persona que, cuando ve cosas rojas, tiene exactamente la misma clase de experiencias que nosotros tenemos cuando vemos cosas verdes, y viceversa; y se nos dice que llamemos a esto un caso de "*qualia* invertidos". Luego se nos pide que imaginemos a una persona aún más peculiar que, al ver rojo, responde a este color de la manera adecuada, que de hecho posee exactamente la misma organización funcional de los estados internos que nosotros, y que, no obstante, no tiene ninguna experiencia en lo absoluto; y se nos dice que llamemos a esto un caso de "*qualia* ausentes". Ahora bien, un partidario de la información fenoménica podría pensar que estas deficiencias tienen algo que ver con lo no físico de la información fenoménica. Pero otros pueden comprenderlas en un sentido diferente. Algunos materialistas rechazarán de plano esos casos, pero otros, entre quienes me cuento, tratarán de entenderlos lo mejor que puedan. Quizá la cuestión radica en que los estados que desempeñan los papeles de la experiencia, y que por ende *son* las experiencias, en las personas normales se encuentran invertidos o ausentes en quienes sufren de *qualia* invertidos o ausentes. (Esto presupone, lo que podría ser falso, que la mayoría de las personas son suficientemente semejantes.) La experiencia de rojo —el estado que desempeña ese papel en la gente normal— también se produce en quien tiene los "*qualia* invertidos", pero en éste desempeña el papel de la experiencia de verde, mientras que el estado que en él desempeña el papel de la experiencia de rojo es el estado que, en las personas normales, desempeña el papel de la experiencia de verde. La experiencia de rojo y de verde —es decir, lo que desempeña esos papeles en las personas normales— no se produce en lo absoluto en quien sufre de "*qualia* ausentes"; lo que en él desempeña esos papeles son estados que no se producen nunca en la gente

[6] Véase Harman 1990, "La cualidad intrínseca de la experiencia" [incluido en esta antología, pp. 263–287].

normal. De esta forma le damos sentido a los *qualia* invertidos y ausentes; pero, en estos términos, "*qualia*" no es sino la palabra para designar los estados que desempeñan papeles tomados *per se*, y no en cuanto que desempeñan papeles. Los *qualia*, así entendidos, no podrían ser asunto de la información fenoménica. María sabe todo acerca de ellos. La mayoría de nosotros, que los tenemos, no sabemos nada.[7]

Es mejor conformarse con un nombre inútil y definido por *vía negativa*. Estipulemos que "el aspecto fenoménico del mundo" nombra cualquier cosa que sea el asunto de la información fenoménica, si tal cosa existe; el aspecto fenoménico, si lo hay, es aquello de lo que podemos obtener información teniendo nuevas experiencias, pero nunca tomando lecciones. Dicho esto, no habrá problema si afirmamos que la información sobre el aspecto fenoménico del mundo sólo puede ser información fenoménica. Pero todo lo que en realidad sabemos, una vez cerrado el círculo de este modo, es que se supone que la información fenoménica revela la presencia de cierta clase de cosas o procesos no físicos dentro de la experiencia, o que se supone que revela que ciertas cosas o ciertos procesos físicos de la experiencia poseen algún tipo de propiedades no físicas.

5. *El argumento del conocimiento*

Si recurrimos a la hipótesis de la información fenoménica a fin de explicar por qué no hay información física suficiente para enseñarnos cómo es una nueva experiencia, tenemos entonces un poderoso argumento para refutar cualquier teoría materialista de la mente. Frank Jackson (véase la nota 1) lo llama el "argumento del conocimiento". Los argumentos para rebatir una u otra teoría materialista nunca son muy concluyentes. Siempre es posible reacomodar los detalles. Pero el argumento del conocimiento, si funcionara, refutaría directamente los postulados mínimos que son comunes a *todas* las teorías materialistas.

Se expresa como sigue. Primero lo expondré en forma simplificada, para luego desarrollarlo apropiadamente. El materialis-

[7] Véanse Block y Fodor, "Lo que no son los estados psicológicos" [incluido en esta antología, pp. 155–181], publicado también en Block 1980; y Lewis 1980, incluido en Block 1980 y en Lewis 1983a.

mo mínimo es una tesis de la superveniencia: no hay diferencia sin diferencia física. En otras palabras, dos posibilidades cualesquiera que son exactamente iguales en términos físicos son exactamente iguales *simpliciter*. Si dos posibilidades son exactamente iguales físicamente, entonces no habrá información física capaz de eliminar una de ellas sin eliminar la otra. Si dos posibilidades son exactamente iguales *simpliciter* (suponiendo que ello sea posible), no habrá entonces ninguna información, del tipo que sea, que pueda eliminar una de ellas sin eliminar la otra. Por lo tanto, si hay una clase de información —a saber, la información fenoménica— capaz de eliminar posibilidades que toda información física deja abiertas, debe haber entonces posibilidades que son exactamente iguales en términos físicos, pero que no son exactamente iguales *simpliciter*. Esto es precisamente lo que niega el materialismo mínimo.

(Analogía. Si dos ubicaciones posibles en nuestra zona coinciden en su coordenada x, no habrá información sobre x que pueda eliminar una sin eliminar la otra. Si, *per impossibile*, dos ubicaciones posibles coinciden en todas sus coordenadas, no habrá ningún tipo de información capaz de eliminar una sin eliminar la otra. Por lo tanto, si hay un tipo de información —a saber, la información y— que puede eliminar ubicaciones que cualquier información x deja abiertas, debe haber entonces ubicaciones en la zona que coinciden con su coordenada x, pero no con todas sus coordenadas.)

Suprimamos ahora la simplificación. Lo que hemos visto hasta ahora es el argumento del conocimiento para refutar el materialismo tomado como una verdad necesaria, que se aplica irrestrictamente a todos los mundos posibles. Pero, por lo general, los materialistas pensamos que el materialismo es una verdad contingente. Concedemos que existen posibles mundos sobrenaturales donde el materialismo es falso, pero insistimos en que nuestro mundo real no se cuenta entre ellos. De ser así, entonces podría haber después de todo dos posibilidades que sean iguales en términos físicos, pero no iguales *simpliciter*, pero una o las dos tendrían que ser posibilidades en un lugar donde el materialismo fuera falso. Los mundos sobrenaturales podrían diferir en cuanto a su sobrenaturalidad sin diferir físicamente. Nuestro materialismo mínimo debe ser una tesis

de la superveniencia *restringida*: dentro de una cierta clase de mundos, que incluye nuestro mundo real, no hay diferencia sin que haya diferencia física. Dentro de esa clase, dos posibilidades cualesquiera que sean exactamente iguales en términos físicos son exactamente iguales *simpliciter*. Pero, ¿qué circunscribe la clase pertinente? (Es una trivialidad decir que nuestro mundo pertenece a *alguna* clase en la cual no hay diferencia sin que haya diferencia física. Esto será así por extraño que sea nuestro mundo. La unidad clase de nuestro mundo, por ejemplo, es una de tales clases. Y lo es cualquier clase que contenga nuestro mundo, y que no contenga dos réplicas físicas.) Creo que la clase pertinente debe ser la que consiste en mundos que no tienen nada totalmente ajeno a este mundo. La población de esos mundos que no nos son ajenos podría estar compuesta por los habitantes del nuestro, por así decirlo, tras un proceso de división y recombinación. Esto daría como resultado que ninguna clase de cosas ni de propiedades fundamentales serían completamente distintas.[8] Nuestra tesis materialista de la superveniencia restringida se expresaría como sigue: en todos los mundos que no nos son ajenos, no hay diferencia sin que haya diferencia física.

Si se admite la hipótesis de la información fenoménica, entonces el argumento del conocimiento refuta esta superveniencia restringida casi tan decididamente como refuta la versión irrestricta. Consideremos una posibilidad que es eliminada por la información fenoménica, pero no por ninguna información física. Tenemos dos casos. Quizá esta posibilidad no contenga nada que sea ajeno a nuestro mundo. En tal caso el argumento se expresa en los mismos términos que el anterior: la realidad y la posibilidad eliminada son exactamente iguales físicamente, no son exactamente iguales *simpliciter*; más aún, ambas caen dentro de la restricción de los mundos que no nos son ajenos, de suerte que tenemos un contraejemplo incluso para la superveniencia restringida. O tal vez, por el contrario, la posibilidad eliminada sí tiene algo *X* que es ajeno a este mundo —una clase de cosa ajena o quizá una propiedad fundamental ajena de las

[8] Véase Lewis 1983b, esp. pp. 361–364. Para una visión diferente sobre cómo postular un materialismo mínimo, véase Horgan 1982.

cosas no ajenas. Entonces, la información fenoménica que se obtiene teniendo una experiencia nueva ha revelado algo negativo: por lo menos en parte, es la información de que *X no* está presente. ¿Cómo puede ser esto? Si la información fenoménica existe, lo que revela es supuestamente positivo: la presencia de algo hasta ahora desconocido. No se trata, desde luego, de algo ajeno a la realidad misma, sino de algo ajeno a la realidad en cuanto que está inadecuadamente representada por los inexpertos y por los materialistas. Si María aprende algo cuando descubre cómo es ver los colores, supuestamente aprende que hay *más* en el mundo de lo que ella conocía hasta entonces, no *menos*. Resulta fácil pensar que la información fenoménica podría eliminar posibilidades que se empobrecen al compararse con la realidad, pero esto sería un contraejemplo a la tesis de la superveniencia restringida. Para eliminar posibilidades sin caer en un contraejemplo, la información fenoménica tendría que eliminar posibilidades menos pobres que la realidad. ¿Y cómo podría lograr esto? Compárese la información perceptual ordinaria. Tal vez Jean-Paul simplemente puede *ver* que Pierre no se encuentra en el café, por lo menos si es un café pequeño. Pero, ¿cómo puede ver que Pierre no se encuentra en París, ya no digamos que no se encuentra en toda la realidad?

(¿Hay un tercer caso? ¿Qué ocurriría si la posibilidad eliminada es, en cierto aspecto, más rica que la realidad, y más pobre en otro aspecto? Supongamos que la posibilidad eliminada tiene *X*, que es ajena a la realidad, pero que a la vez le falta *Y*. La información fenoménica podría entonces eliminarla si pone de manifiesto la presencia real de *Y*, sin necesidad de poner de manifiesto la ausencia real de *X*. Pero entonces digo que debería haber una tercera posibilidad, la cual no tuviera ni *X* ni *Y*, más pobre y no más rica en ninguno de sus aspectos que la realidad, y otra vez sin ninguna diferencia física con respecto a la realidad. Pues, ¿por qué el hecho de suprimir *X* reestablece automáticamente a *Y*? ¿Por qué no pueden variar independientemente?[9] Pero esta tercera posibilidad difiere *simpliciter* de la

[9] Sobre la recombinación de posibilidades, véase mi 1986a, pp. 87-92. El presente argumento podría requerir un principio que también permita la recombinación de propiedades; ahora pienso que esto no necesariamen-

realidad sin diferir en el nivel físico. Aún más, no tiene nada ajeno a la realidad. Así, volvemos a obtener un contraejemplo de la tesis de la superveniencia restringida.)

El argumento del conocimiento funciona. No hay manera de dar por sentada la hipótesis de la información fenoménica y seguir sosteniendo el materialismo. En consecuencia, niego la hipótesis. No puedo refutarla en forma directa. Sin embargo, más adelante argumentaré, primero, que es más peculiar, y por ende menos tentadora, de lo que al principio podría parecer; y, en segundo lugar, que no estamos obligados a aceptarla, ya que una hipótesis alternativa le hace justicia al modo en que la experiencia enseña mejor el cómo es.

6. *Otras tres maneras de no captar la idea*

La hipótesis de la información fenoménica caracteriza la información en términos de posibilidades eliminadas. Pero hay otras concepciones de "información". En consecuencia, la hipótesis posee variantes semejantes: hipótesis que postulan que la experiencia produce "información" que no puede obtenerse de otro modo, pero que no caracterizan esta "información" en términos de posibilidades eliminadas. Estas variantes no sirven como premisas del argumento del conocimiento. No dicen que la información fenoménica elimina posibilidades que difieren, aunque no físicamente, de posibilidades no eliminadas. Las hipótesis semejantes de la "información" fenoménica son consistentes con el materialismo, y bien pueden ser verdaderas. Pero no suprimen el argumento del conocimiento. Independientemente de cuáles variantes inofensivas puedan o no ser verdaderas, e independientemente de qué concepción pueda o no merecer el nombre de "información", la única forma de salvar al materialismo es concentrar de lleno nuestra atención en la genuina hipótesis de la información fenoménica, y negarla. Apartar la mirada y prestar atención a otra cosa no servirá como sustituto de esa negación.

te requeriría tratar las propiedades como partes no espacio-temporales de sus instancias. Acerca de la recombinación de propiedades, véase también Armstrong 1989.

¿Podría una de esas variantes ayudarnos siquiera en el sentido siguiente: darnos algo verdadero que bien podría haberse confundido con la hipótesis genuina, explicando así por qué hemos sido llevados a creer en la hipótesis, aunque sea falsa? Creo que no. Resulta que cada una de las variantes implica no sólo que la experiencia puede proporcionarnos "información" que no podemos obtener a través de lecciones, por muchas que sean, sino también que las lecciones en ruso pueden darnos "información" que no podemos obtener con las lecciones en inglés (y viceversa). Dudo que un partidario de la información fenoménica pensara jamás que el papel especial de la experiencia al enseñarnos cómo es ser algo estaría en el mismo nivel que el del papel especial del ruso. Antes de terminar tendré que decir que la información fenoménica es una ilusión, pero creo que debo buscar en otra parte para hallar una hipótesis creíble sobre el tipo de ilusión que podría ser.

La cuarta manera. Si una cámara oculta toma fotografías de un cuarto, la película acabará por contener rastros de lo que ocurrió en ese cuarto. Los rastros son distintivos, esto es, los detalles de esos rastros dependen de los detalles de lo que ocurrió, y si lo que ocurrió hubiera sido diferente en cualquiera de muchos aspectos, los rastros habrían resultado correspondientemente diferentes. Así, podemos decir que los rastros tienen información, y que la persona que posee la película tiene la información. Podría decirse esto porque los rastros, aunados a la manera en que dependen de lo que ocurrió, bastan para eliminar posibilidades; pero, en cambio, podríamos decir "información", y significar únicamente "rastros distintivos". De ser así, es indudablemente verdadero que la nueva experiencia transmite "información" distinta de cualquier otra que pueda obtenerse con unas lecciones. La experiencia y las lecciones dejan diferentes clases de rastros. Esto es así independientemente de que la experiencia elimine posibilidades que las lecciones dejan abiertas. Es igualmente verdadero, desde luego, que las lecciones en ruso dejan rastros distintos de los que dejan las lecciones en inglés, al margen de si las lecciones cubren o no la misma materia y eliminan las mismas posibilidades.

La quinta manera. Cuando hablamos de transmisión de "información", a menudo queremos decir transmisión de un texto.

Los depósitos de "información", como las bibliotecas, son almacenes de textos. Que el texto sea una verborrea vacua o sumamente informativo no viene al caso. Quizá también nosotros contengamos información como almacenes de texto. Quizá haya un lenguaje del pensamiento y quizá la manera en que creemos las cosas consista en almacenar enunciados en este lenguaje de un modo especial o en una sección especial de nuestro cerebro. En tal caso, podríamos decir que almacenar una nueva oración es almacenar una nueva "información", independientemente de que ésta elimine posibilidades que no han sido ya eliminadas por las oraciones almacenadas previamente. Quizá, también, el lenguaje del pensamiento no se encuentre establecido de una vez y para siempre, sino que puede adquirir nuevas palabras. Quizá, por ejemplo, tome palabras prestadas del lenguaje público. Y quizá, cuando alguien vive una nueva experiencia, esto ocasione que su lenguaje del pensamiento adquiera una nueva palabra que denota esa experiencia, una palabra que no se podría haber añadido al lenguaje por otros medios. Si todo esto es así, entonces cuando María ve colores, su lenguaje del pensamiento adquiere nuevas palabras, permitiéndole almacenar nuevas oraciones y, por ende, añadir "información". Todo lo anterior, acerca del lenguaje del pensamiento, el almacenaje de oraciones y la adquisición de palabras es una especulación. Pero es una especulación plausible, aunque no la única que se maneja. Si todo ello es verdadero, tenemos entonces otra hipótesis semejante de la "información" fenoménica. Cuando María adquiere nuevas palabras y almacena nuevas oraciones, ésta es "información" que nunca había tenido antes, independientemente de si esto elimina posibilidades que María no hubiera ya eliminado.

Pero también en este caso el papel especial de la experiencia resulta estar a la par con el papel especial del ruso. Si el lenguaje del pensamiento adquiere nuevas palabras, tomándolas prestadas del lenguaje público, las lecciones en ruso añaden entonces nuevas palabras y tienen como resultado el almacenaje de nuevas oraciones, inculcando así "información" que jamás se podría haber obtenido con las lecciones en inglés. (Cabría objetar que las nuevas palabras rusas no son más que sinónimos de palabras viejas, o por lo menos de frases viejas, que ya esta-

ban allí, y que los sinónimos no cuentan. Pero no se ha ofrecido razón alguna para que las nuevas palabras internas creadas por la experiencia no puedan ser también sinónimos de viejas frases, tal vez de largas descripciones formuladas en el lenguaje de la neurofisiología.)

La sexta manera. Un filósofo escéptico acerca de posibilidades, como es el caso de muchos, tal vez quiera reemplazar las posibilidades mismas por posibilidades lingüísticas sustitutas: conjuntos máximos consistentes de oraciones. Y podría conformarse con interpretar "consistente" en un sentido estrictamente lógico, de suerte que un conjunto formado por "Alfredo está casado" y "Alfredo es soltero" cuente como consistente, y sólo sean descartadas las contradicciones abiertas, como "Alfredo está casado" y "Alfredo no está casado".[10] También se podrían considerar las posibilidades sustitutas o simuladas como conjuntos de oraciones en el lenguaje del pensamiento, si el filósofo cree en éste. Así, si el lenguaje del pensamiento de una persona adquiere nuevas palabras, ya sea como resultado de una nueva experiencia o como resultado de haber tomado lecciones en ruso, las posibilidades simuladas se vuelven más ricas y numerosas. Los conjuntos de oraciones que antes eran máximos dejan de serlo en cuanto se añaden nuevas palabras. De este modo, cuando María ve los colores y su lenguaje del pensamiento adquiere nuevas palabras, aparecen nuevas posibilidades simuladas, y ella puede eliminar algunas en el acto. Supongamos que sabe de antemano que está a punto de ver verde y que la experiencia de ver verde está relacionada con el patrón *F* de activación neuronal. Así, cuando ve verde y adquiere la nueva palabra *G* para designar su experiencia, se producen de inmediato posibilidades simuladas nuevas y enriquecidas, las cuales incluyen oraciones que dicen que María tiene *G* sin *F*, y en el acto ella sabe lo suficiente para eliminar estas posibilidades simuladas. (Incluso, si María no sabe de antemano lo que está a punto de ver, puede eliminar de inmediato por lo menos aquellas posibilidades simuladas recientemente descubiertas cuyas oraciones nieguen que María tiene ahora *G*.) Tal

[10] Véase Lewis 1986a, pp. 142–165, con respecto a las posibilidades lingüísticas sustitutas.

como podemos caracterizar la información en términos de la eliminación de posibilidades, también podemos caracterizar la "información" simulada en términos de la eliminación de "posibilidades" simuladas. De modo que aquí tenemos la variante más fielmente parecida a la hipótesis, siempre y cuando el lenguaje del pensamiento sea verdadero. Pero todavía no tenemos la genuina hipótesis de la información fenoménica, porque la posibilidad simulada eliminada acerca de *G* sin *F* puede no haber sido en lo absoluto una posibilidad genuina. Quizá fue algo parecido a la posibilidad sustituta de los solteros casados.

7. *Curioso y cada vez más curioso*

La hipótesis de la información fenoménica es más peculiar de lo que puede parecer a primera vista. Por un lado, porque no sólo se opone al materialismo. Algunos de ustedes quizá hayan acogido de buen grado el argumento del conocimiento debido a que siempre pensaron que la información física era inadecuada para explicar los fenómenos de la mente. Tal vez ustedes siempre estuvieron convencidos de que la mente podía hacer cosas que ningún sistema físico puede hacer: doblar cucharas, inventar chistes, demostrar la consistencia de la aritmética, reducir el paquete de ondas, o lo que sea. Quizá ustedes estaban convencidos de que toda la historia causal de la manera en que se realizan los actos de la mente implica las interacciones causales no sólo de los cuerpos materiales, sino también de los cuerpos astrales; no sólo de las vibraciones del campo electromagnético, sino también de las buenas o malas vibras del campo psiónico; no sólo del protoplasma, sino también del ectoplasma. Yo tengo mis dudas, pero no importa. Es irrelevante para nuestro tema. El argumento del conocimiento está dirigido contra ustedes, tanto como contra el propio materialismo.

Digamos que la *parapsicología* es la ciencia de todas las cosas, las propiedades, los procesos causales, las leyes de la naturaleza y demás, que no son físicos y que pueden necesitarse para explicar lo que hacemos. Supongamos que llegamos a saber todo de la parapsicología. No producirá diferencia alguna. María, en su mundo en blanco y negro, puede estudiar todo sobre la parapsicología, así como todo sobre la psicofísica de la visión

en colores, pero seguirá sin saber cómo es. Las lecciones sobre el aura de Vegemite no nos ayudarán más que las lecciones sobre su composición química, y así con otras cosas. Nuestro punto de partida intuitivo no era sólo que las lecciones de *física* no podrían ayudar a quienes no han vivido una experiencia de cómo es ser algo. Era que las *lecciones* no podían ayudar. Si la información fenoménica existe, no es sólo independiente de la información física. Es independiente de todo tipo de información que pueda presentarse en las lecciones a quienes no han vivido la experiencia. Pues se supone que elimina posibilidades que cualquier cantidad de lecciones dejan abiertas. En consecuencia, la información fenoménica no es sólo información parapsicológica, si tal cosa existe. Es algo mucho más extraño.

La genuina hipótesis de la información fenoménica, en tanto se distingue de sus variantes, trata la información en términos de la eliminación de posibilidades. Cuando carecemos de información, están abiertas varias posibilidades alternativas; cuando obtenemos la información, algunas de las alternativas quedan excluidas. Pero una segunda peculiaridad de la información fenoménica es que se resiste a este tratamiento. (Como también la "información" lógica o matemática. Sin embargo, la información fenoménica no puede ser lógica o matemática, porque las lecciones de lógica o matemática no nos enseñan mejor cómo es una nueva experiencia que las lecciones de física o parapsicología.) Cuando alguien no sabe cómo es tener una experiencia, ¿dónde están las posibilidades alternativas que quedan abiertas? No puedo representarme en la mente una gama de posibilidades alternativas acerca de cómo podría ser probar Vegemite. Esto se debe a que no puedo imaginar ni cómo *es* probar Vegemite, ni ninguna otra forma alternativa de cómo *podría* ser, pero que de hecho, no es. (Yo podría imaginar perfectamente que Vegemite sabe igual que la mantequilla de cacahuate, o alguna otra cosa que me es familiar, pero supongamos que un conocedor me ha asegurado que no es así.) Ni siquiera puedo plantear la pregunta que supuestamente debe responder la información fenoménica: ¿es de ésta o de aquella manera? Parecería que las alternativas posibles deben ser impensables de antemano; y posteriormente también, salvo para la que al final

de cuentas se ha realizado. No estoy diciendo que haya algo completamente imposible acerca de una gama de alternativas impensables; sólo que hay algo peculiar. Pero es ya bastante peculiar sugerir que de algún modo pudimos haber errado el camino.

8. *De lo fenoménico a lo epifenoménico*

Una tercera peculiaridad de la información fenoménica es que se encuentra extrañamente aislada de todas las demás clases de información; y esto es así independientemente de que la mente opere con base en principios físicos o parapsicológicos. El aspecto fenoménico del mundo no tiene nada que ver con la explicación de por qué las personas aparentemente hablan del aspecto fenoménico del mundo. Por ejemplo, no desempeña ningún papel en la explicación de los movimientos de los bolígrafos que usan los filósofos al escribir sus tratados sobre información fenoménica ni de la manera en que la experiencia los ha provisto de esta última.

Cuando María sale de su habitación en blanco y negro, se queda boquiabierta. "¡Por fin!", exclama. "¡De modo que esto es lo que es ver colores!" Después hace cosas que no podía hacer antes, como reconocer una nueva muestra del primer color que vio en su vida. También puede hacer otras cosas que no hacía antes: cosas desafortunadas, como escribir sobre la información fenoménica y la pobreza del materialismo. Podría pensarse que María dijo lo que dijo e hizo lo que hizo porque llegó a conocer cómo es ver los colores. No es así, si la hipótesis de la información fenoménica es correcta. Pues supongamos que el aspecto fenoménico del mundo hubiera sido distinto, de suerte que María adquiriera una información fenoménica diferente. O supongamos que no hubiera habido ningún aspecto fenoménico del mundo, como pensamos los materialistas. ¿Habría significado la más mínima diferencia en lo que María hizo o dijo en ese momento o posteriormente? Creo que no. Una diferencia en lo que ella hace o dice significa, por lo menos en parte, una diferencia en los movimientos de las partículas que componen a María. (O mejor: una diferencia en la forma espaciotemporal de la función de onda de esas partículas. Pero dejémoslo así.) Pues,

¿cómo podría ella hacer o decir algo diferente, si ninguna de sus partículas se movió de manera diferente? Pero si algo que no es físico produce a veces una diferencia en los movimientos de las partículas físicas, entonces la física, tal como la conocemos, es errónea. No sólo peca de omisión, no sólo es incompleta, sino errónea. O bien algo ocasiona que las partículas cambien de movimiento sin la intervención de ninguna fuerza, o bien hay alguna fuerza adicional que opera de manera muy distinta de las cuatro habituales. Creer en el aspecto fenoménico del mundo, pero negar que es epifenoménico, es apostar en contra de la verdad de la física. Dado el éxito que hasta ahora ha tenido la física, e incluso con las debidas concesiones a los males fundacionales de la mecánica cuántica, ésta sería una apuesta precipitada. Un partidario del aspecto fenoménico estaría más seguro aliándose con Jackson en la defensa de los *qualia epifenoménicos*.

Pero hay más en juego que una simple apuesta empírica a favor de la física. Supongamos que hay un aspecto fenoménico del mundo, y supongamos que sí produce una diferencia en los movimientos de la mandíbula de María o en los sonidos que salen de su boca. Podemos entonces describir el aspecto fenoménico en términos de sus efectos físicos, si tenemos suficiente conocimiento. Es de esto de lo que dependen de tal y cual manera los fenómenos físicos. Este asidero descriptivo nos permitirá dar lecciones sobre el tema a quienes no han vivido la experiencia. Pero en la medida en que podemos dar lecciones sobre el tema, lo que tenemos es sólo parapsicología. Aquello que no podemos aprender si no es teniendo la experiencia sigue eludiéndonos. No estoy argumentando que *todo* lo relativo al supuesto asunto de la información fenoménica tenga que ser epifenoménico. Parte de ello podría ser, en cambio, parapsicológico. Pero insisto en que *alguno* de sus aspectos tiene que ser epifenoménico.

Supongamos que la hipótesis de la información fenoménica es verdadera y supongamos que V_1 y V_2 son todas las posibilidades fenoménicas máximamente específicas referentes a cómo es probar Vegemite; quien pruebe Vegemite descubrirá cuál de las dos se produce, y nadie más puede descubrir esto. Y supongamos que P_1 y P_2 son todas las posibilidades físicas máximamente

específicas. (Desde luego, necesitamos en realidad mucho más que dos P, y quizá un partidario de la información fenoménica querría más de dos V, pero un conjunto absurdamente pequeño servirá para el ejemplo.) Tenemos entonces cuatro hipótesis alternativas sobre la independencia o dependencia causal de las P con relación a las V. Cada una puede expresarse como un par de condicionales contrafácticos. Hay dos hipótesis que son patrones de dependencia.

K_1: si V_1 entonces P_1, si V_2 entonces P_2
K_2: si V_1 entonces P_2, si V_2 entonces P_1

Las otras dos son patrones de independencia.

K_3: si V_1 entonces P_1, si V_2 entonces P_1
K_4: si V_1 entonces P_2, si V_2 entonces P_2

Estas hipótesis de la dependencia son, según entiendo, proposiciones contingentes. Las hace verdaderas, si lo son, algún rasgo contingente del mundo, aunque qué tipo de rasgo sea es un asunto fastidioso de decidir.[11] Ahora tenemos ocho posibilidades conjuntas.

K_1 V_1 P_1	K_3 V_1 P_1	K_3 V_2 P_1	K_2 V_2 P_1
K_2 V_1 P_2	K_4 V_1 P_2	K_4 V_2 P_2	K_1 V_2 P_2

Entre las cuatro posibilidades de la hilera superior y las cuatro de la inferior está la diferencia física entre P_1 y P_2. Entre las cuatro de la izquierda y las cuatro de la derecha está la diferencia fenoménica entre V_1 y V_2. Y entre las cuatro de los costados y las cuatro del medio hay una diferencia parapsicológica. Es la diferencia entre la dependencia y la independencia de lo físico con respecto a lo fenoménico; entre la eficacia y el epifenomenalismo, en lo que concierne a este ejemplo. Esto no tiene nada de inefable. Ya sea que hayas probado o no Vegemite, y que puedas o no concebir la presunta diferencia entre V_1 y V_2, todavía te pueden decir si la diferencia física entre P_1 y P_2 depende o no de cierta parte del aspecto fenoménico del mundo.

Las lecciones pueden enseñar a quienes no han vivido la experiencia cuál es la posibilidad parapsicológica que se produce,

[11] Sobre la hipótesis de la dependencia, véase Lewis 1981.

la de la dependencia o la de la independencia. Supongamos que se trata de la posibilidad de la dependencia: tenemos o bien K_1 o bien K_2. Pues si tuviéramos independencia, habríamos ya descubierto nuestra diferencia epifenoménica, a saber, la diferencia entre V_1 y V_2. Y las lecciones pueden enseñar a quienes no han vivido la experiencia cuál de las dos posibilidades físicas se da. Sin perder la generalidad, supongamos que es P_1. Ahora bien, dos de nuestras ocho posibilidades conjuntas originales siguen abiertas: K_1 V_1 P_1 y K_2 V_2 P_1. La diferencia entre ellas no es en absoluto física ni tampoco parapsicológica: es P_1 y su dependencia, en ambos casos. La diferencia es enteramente fenoménica. Y también es enteramente epifenoménica. No hay nada físico, ni nada parapsicológico, que dependa de la diferencia entre K_1 V_1 P_1 y K_2 V_2 P_1. Tenemos el mismo tipo de patrón de dependencia en ambos sentidos; sólo ocurre que las posibilidades fenoménicas han sido intercambiadas. Por tanto, ya se trate de la independencia o de la dependencia, hemos encontrado una parte epifenoménica en el aspecto fenoménico del mundo. Es el residuo que queda cuando suprimimos la parte parapsicológica.

Supongamos que un día pruebo la Vegemite, y expongo cómo por fin llegué a saber cómo sabe. El sonido de mi exposición es un efecto físico, parte de la posibilidad física realizada P_1. Este efecto físico es exactamente el mismo, ya forme parte de la posibilidad conjunta K_1 V_1 P_1 o de su alternativa K_2 V_2 P_1. Puede ser causada por V_1 en concordancia con K_1 o, por el contrario, puede ser causada por V_2 en concordancia con K_2, pero en uno u otro caso es lo mismo. Así, no ocurre debido a que tenemos K_1 V_1 en vez de K_2 V_2, o viceversa. La supuesta diferencia entre estas dos posibilidades no ayuda en nada a explicar la presunta manifestación física de que yo descubra cuál de ellas se ha realizado. Es de esta manera como la diferencia resulta epifenoménica. Esto lo hace muy extraño, y repugnante a la sensatez.

9. *La hipótesis de la habilidad*

De este modo, la hipótesis de la información fenoménica resulta ser en realidad muy peculiar. Sería bueno, y no sólo para los

materialistas, que pudiéramos refutarla. Para los materialistas, rechazarla es esencial. Y podemos hacerlo. Hay una hipótesis alternativa acerca de lo que es aprender cómo es una experiencia: la *hipótesis de la habilidad*. Laurence Nemirow la resume en estos términos:

> algunas formas de entendimiento no consisten en asimilar hechos, sino en la adquisición de habilidades [...]. En lo que respecta al entendimiento de una experiencia, podemos interpretarlo como la habilidad de colocarse, a voluntad, en un estado representativo de esa experiencia. Entiendo la experiencia de ver rojo si soy capaz de visualizar el rojo a voluntad. Ahora bien, es perfectamente claro por qué debe haber una conexión especial entre la habilidad de colocarse en un estado representativo de una experiencia dada y el punto de vista del experimentador: ejercer esa habilidad es sencillamente lo que llamamos "asumir el punto de vista del experimentador". [...] Podemos llegar a comprender, pues, la subjetividad de nuestro entendimiento de la experiencia sin postular hechos subjetivos como objetos de nuestro entendimiento. Esta caracterización explica, incidentalmente, la imposibilidad de comunicar lingüísticamente nuestro entendimiento subjetivo de la experiencia (un fenómeno que, según podría parecer, apoya la hipótesis de los hechos subjetivos). Esto último se explica como un caso especial de la imposibilidad de comunicar lingüísticamente las habilidades que se necesitan para colocarse a voluntad en un estado dado, como el estado de haber logrado bajar la presión arterial y el estado de agitar las orejas.[12]

Si tienes una nueva experiencia, adquieres habilidades para recordar e imaginar. Una vez que has probado la Vegemite, y aprendiste cómo es, puedes recordar más tarde la experiencia vivida. Al recordar cómo fue en esa ocasión, puedes imaginar después una experiencia semejante. De hecho, incluso si acabas

[12] Nemirow 1980, pp. 475–476. Para una exposición más completa, véase Nemirow 1989 y 1979. Véase también Tye 1986.

Debo dejar sentado un desacuerdo con Nemirow acerca de una pequeñísima cuestión. Coincidimos en que la frase "cómo es la experiencia *E*" no denota una "cualidad subjetiva" de *E*, algo que supuestamente formaría parte del tema de la información fenoménica adquirida al tener *E*. Pero mientras que yo he interpretado que la frase denota a la *E* misma, Nemirow la interpreta como una parte sincategoremática de la expresión "saber cómo es la experiencia *E*". Véase Nemirow 1989, sec. III.

por olvidar la ocasión misma, es muy probable que conserves la habilidad de imaginar esa experiencia.

Aún más, adquieres la habilidad de reconocer la misma experiencia si vuelve a producirse. Si pruebas la Vegemite otro día, probablemente sabrás que has paladeado ese sabor antes. Y si, mientras la pruebas, sabes que es Vegemite lo que estás probando, podrás ponerle nombre a la experiencia si la vuelves a tener. O si no te dicen nada en el momento, pero más adelante te enteras de algún modo de que eso cuyo sabor estás entonces recordando o imaginando es Vegemite, podrás también ponerle nombre a la experiencia, o al recuerdo, o a la experiencia de imaginar, si vuelve a producirse. En este caso, la habilidad adquirida es la habilidad de adquirir información si recibes nueva información. Sin embargo, la información adquirida no es fenoménica, y la habilidad de adquirir información no es lo mismo que la información misma.

Anteriormente hablé de "saber cómo es una experiencia bajo cierta descripción". Ahora puedo decir que lo que pretendía significar con esto era tener la habilidad de recordar o imaginar una experiencia y conocer, al mismo tiempo, la proposición egocéntrica de que aquello que uno está imaginando en ese momento es la experiencia de tal o cual descripción. Uno bien puede saber cómo es una experiencia bajo cierta descripción, pero no bajo otra. Incluso puede saber cómo es una experiencia, pero no bajo ninguna descripción, cualquiera que ésta sea, a menos que se trate de una descripción bastante trivial, como "ese extraño sabor que estoy imaginando en este instante". Esto es lo que sucedería si alguien pusiera una pizca de Vegemite en mi comida sin decirme lo que es; luego yo sabría cómo es probar la Vegemite, pero no bajo esa descripción, ni bajo ninguna otra descripción no trivial. Podría sugerirse que "saber cómo es probar la Vegemite" significa en realidad lo que yo llamaría "saber cómo es probar la Vegemite bajo la descripción 'probar Vegemite' "; y de ser así, saber cómo es implicaría tanto habilidad como información. No estoy de acuerdo; pues sin duda tendría sentido decir: "Conozco bien esta experiencia, sé desde hace mucho cómo es, pero apenas hoy he descubierto que es la experiencia de probar la Vegemite." Pero esta cuestión verbal carece de importancia, ya que la información implicada en

saber cómo es bajo cierta descripción, y presuntamente implicada en saber cómo es, no constituye de todos modos la extraña información fenoménica que debe ser rechazada.

(¿Habrá en esto algún problema para el partidario de la información fenoménica? Supongamos que éste dice que saber cómo es probar la Vegemite significa saber que el sabor de la Vegemite posee cierto "carácter fenoménico". Esto exige ponerle nombre al sabor, de manera que corresponda claramente con nuestra noción de saber cómo es probar la Vegemite bajo la descripción "probar Vegemite". Pero también tenemos nuestra noción de saber cómo es *simpliciter*, ¿y qué puede él ofrecernos que corresponda con esto? Quizá deba responder recurriendo a una descripción trivial, en los términos siguientes: saber cómo es *simpliciter* significa saber cómo es bajo la descripción trivial "el sabor que ahora estoy imaginando", lo cual implica saber que el sabor que uno ahora está imaginando posee cierto carácter fenoménico.)

Además de adquirir la habilidad de recordar y de imaginar la experiencia que tuviste, también adquieres la habilidad de imaginar experiencias afines que nunca has tenido. Después de probar Vegemite, podrías desarrollar la habilidad, por ejemplo, de imaginar el sabor de un helado de Vegemite. Realizando experimentos imaginativos, puedes predecir con cierta seguridad lo que harías en circunstancias que no se han presentado, por ejemplo, si pidieras una segunda porción de helado de Vegemite.

Estas habilidades de recordar, imaginar y reconocer no se pueden adquirir (salvo mediante una superneurocirugía o por un acto de magia) a menos que hayas probado la Vegemite y hayas aprendido cómo es. No se las puede obtener tomando lecciones sobre la física o la parapsicología de la experiencia, ni siquiera tomando lecciones exhaustivas que abarquen toda la física y la parapsicología. La hipótesis de la habilidad dice que saber cómo es una experiencia *es* simplemente la posesión de esas habilidades de recordar, imaginar y reconocer. No es la posesión de ninguna clase de información, sea ordinaria o peculiar. No es el conocimiento de que ciertas posibilidades no están realizadas. No es saber que. Es saber cómo. En consecuencia, no debería sorprendernos que las lecciones no enseñen cómo

es una experiencia. Las lecciones inculcan información; la habilidad es otra cosa. Saber que no es automáticamente saber cómo.

Hay casos paralelos. Algunos saben cómo mover las orejas; otros no. Si alguien no puede hacerlo, no habrá información capaz de ayudarlo. Algunos saben comer con palillos, otros no. La información es útil hasta cierto punto –por ejemplo, si el problema consiste en que uno sostiene un palillo en cada mano–, pero no habrá información que, por sí misma, le permita a uno alcanzar un alto grado de saber cómo. Algunos saben cómo reconocer una locomotora C-38 a simple vista, otros no. Si alguien no sabe hacerlo, no lo ayudará mucho memorizar una detallada descripción geométrica de su forma, aun cuando esto permita la eliminación de todas las posibilidades que deben ser eliminadas. (A la inversa, conocer la forma mediante la vista no le confiere la habilidad de dar, por escrito, su descripción geométrica.) La información contribuye muy frecuentemente a saber-cómo, pero a menudo no contribuye en grado suficiente. Ésta es la razón por la que los estudiantes de música deben practicar.

Saber cómo es una habilidad. Pero, desde luego, algunos aspectos de la habilidad no son conocimiento en ningún sentido: fuerza, dinero suficiente. Otros aspectos de la habilidad son, lisa y llanamente, una cuestión de información. Si uno quiere saber cómo abrir la puerta de combinación de la bóveda de un banco, todo lo que necesita es información. Queda el hecho de que hay aspectos de la habilidad que *no* consisten en la mera posesión de información, pero que *sí* llamamos conocimiento. La hipótesis de la habilidad sostiene que saber cómo es una experiencia consiste en ese tipo de conocimiento.

Si la hipótesis de la habilidad es la manera correcta de analizar el saber cómo es una experiencia, entonces la información fenoménica es una ilusión. Deberíamos explicar esa ilusión. Sería endeble, creo, decir que nos dejamos engañar por la ambigüedad de la palabra "saber": confundimos la habilidad con la información porque confundimos el conocimiento en el sentido de saber cómo con el conocimiento en el sentido de saber que. Quizá la palabra "saber" tenga dos sentidos, pero están

perfecta y genuinamente entremezclados. Marcan los dos extremos puros de un espectro de casos mixtos. Lo habitual es que adquiramos información y habilidad al mismo tiempo. De ser así, no deberíamos sorprendernos si aplicamos a los casos puros de adquisición de una habilidad o a los casos puros de adquisición de información el mismo término "saber" que aplicamos a todos los casos mixtos.

Junto con la información y la habilidad, la familiaridad es un tercer elemento de la mezcla. Si Lloyd George murió a edad muy temprana, hay cierto sentido en el que mi padre nunca lo conocerá. La información no le ayudará, aunque mi padre sea un biógrafo muy puntilloso y aunque los archivos estén perfectamente completos. (Y el problema no radica en que haya información muy especial acerca de una persona que sólo pueda obtenerse estando en su presencia.) Saber cómo tampoco ayudará, por mucho que mi padre pueda imaginarse a Lloyd George, en apariencia recordándolo y reconociéndolo. (Mi padre puede ser capaz de reconocer a Lloyd George aunque ya no exista un Lloyd George a quien reconocer —si *per impossibile* éste apareciera, mi padre podría decir que se trata de él.) Nuevamente, lo que tenemos no es sólo un tercer sentido independiente de "saber". Encontrarte con alguien, adquirir mucha información acerca de él que sería difícil conseguir de otra manera y adquirir habilidades respecto de él suelen ser cosas que van juntas. Los casos puros son excepciones.

Un partidario de la información fenoménica estará de acuerdo, desde luego, en que cuando aprendemos cómo es una experiencia, adquirimos las habilidades de recordar, imaginar y reconocer. Pero dirá que adquirimos esas habilidades porque obtenemos información fenoménica. Podría decir incluso lo mismo con respecto a otros casos en que se llega a saber-cómo: uno puede reconocer la locomotora C-38 cuando cuenta con información fenoménica acerca de cómo es ver esa forma, uno puede comer con palillos o mover las orejas cuando obtiene información fenoménica sobre la experiencia de hacer esas cosas, y así en otros casos. ¿Qué debemos entender con esto los partidarios de la hipótesis de la habilidad? ¿Está aquél ofreciendo una conjetura, que debemos rechazar, sobre el origen

causal de las habilidades? Creo que no. Piensa, como nosotros, que las experiencias dejan rastros distintivos en las personas, y que estos rasgos nos permiten hacer cosas. De modo semejante, el hecho de que nos enseñen a reconocer una locomotora C-38 o a comer con palillos, o lo que sucede cuando se mueven las orejas por primera vez, deja rastros que posteriormente nos permiten hacer cosas. Hasta aquí estamos en un campo común. El partidario de la información fenoménica también interpreta estos rastros habilitadores como representaciones que conllevan información sobre sus causas. (Si los mismos rastros hubieran sido causados de una forma equívoca, podrían quizá haber transmitido información errónea.) Podríamos incluso aceptar esto también. El momento de la polémica sólo se produce cuando afirma que estos rastros representan hechos fenoménicos especiales, hechos que no se pueden representar de ninguna otra forma y que, por lo tanto, no pueden ser enseñados en las lecciones de física o incluso en las lecciones de parapsicología. Ésta es la parte, y la *única* parte, que debemos rechazar. Pero no es una parte de su explicación psicológica sobre la manera en que adquirimos habilidades. Sólo se trata de una glosa metafísica gratuita de esa explicación.

Decimos que aprender cómo es una experiencia significa adquirir ciertas habilidades. Si la base causal de esas habilidades resulta ser también una clase especial de representación de algún tipo de información, no hay problema. Sólo necesitamos negar que representa un tipo especial de información sobre un tema especial. Aparte de esto, está abierta la cuestión de qué representaría, si acaso representa algo. ¿Los pormenores de los estímulos –la composición química de la Vegemite, las reflexiones en las superficies, los movimientos de unos palillos diestramente manipulados o de las orejas? ¿Los pormenores de los estados internos producidos por esos estímulos –los patrones de las activaciones nerviosas? Podríamos estar de acuerdo con ambas cosas, siempre que no confundamos "tener información" representada de esta manera especial con tener la misma información bajo la forma de conocimiento o creencia. O podríamos no estar de acuerdo. Tratar el rastro que confiere habilidades como una representación es algo opcional. Lo esencial

radica en que cuando aprendemos cómo es una experiencia teniendo esa experiencia, adquirimos las habilidades de recordar, imaginar y reconocer.

[*Traducción de Ana Isabel Stellino*]

BIBLIOGRAFÍA

Armstrong, D.M., 1989, *A Combinatorial Theory of Possibility*, Cambridge University Press, Cambridge.

Block, N., 1980 (comp.), *Readings in Philosophy of Psychology*, vol. 1, Harvard University Press, Cambridge, Mass.

——–, y J.A. Fodor, 1972, "What Psychological States Are Not", *Philosophical Review*, vol. 81, pp. 159–181. [Incluido en esta antología, pp. 155–181.]

Chisholm, R., 1981, *The First Person: An Essay on Reference and Intentionality*, University of Minnesota Press, Mineápolis.

Churchland, P.M., 1985, "Reduction, Qualia, and the Direct Introspection of Brain States", *Journal of Philosophy*, vol. 82, pp. 8–28.

Farrell, B.A., 1950, "Experience", *Mind*, vol. 59, pp. 170–198.

Harman, G., 1990, "The Intrinsic Quality of Experience", *Philosophical Perspectives*, vol. 4, pp. 31–52. [Incluido en esta antología, pp. 263–287.]

Horgan, T., 1982, "Supervenience and Microphysics", *Pacific Philosophical Quarterly*, vol. 63, pp. 29–43.

Jackson, F., 1982, "Epiphenomenal Qualia", *Philosophical Quarterly*, vol. 32, no. 127, pp. 127–136. [Incluido en esta antología, pp. 95–110.]

———, 1986, "What Mary Didn't Know", *Journal of Philosophy*, vol. 83, pp. 291–295. [Incluido en esta antología, pp. 111–118.]

Lewis, D., 1986a, *On the Plurality of Worlds*, Blackwell, Oxford.

———, 1986b, *Philosophical Papers*, vol. II, Oxford University Press, Nueva York.

———, 1983a, *Philosophical Papers*, vol. I, Oxford University Press, Nueva York.

———, 1983b, "New Work for a Theory of Universals", *Australasian Journal of Philosophy*, vol. 61, pp. 343–377.

———, 1980, "Mad Pain and Martian Pain", en Block 1980, pp. 216–222; y en Lewis 1983a, pp. 122-130.

Lewis, D., 1981, "Causal Decision Theory", *Australasian Journal of Philosophy*, vol. 59, pp. 5–30; reimpreso en Lewis 1986b, pp. 305–339.
———, 1979, "Attitudes *De Dicto* and *De Se*", *Philosophical Review*, vol. 88, pp. 513–543.
Lycan, W.G. (comp.), 1980, *Mind and Cognition: A Reader*, Basil Blackwell, Oxford.
Nagel, T., 1974, "What Is It Like to Be a Bat?", *Philosophical Review*, vol. 83, no. 4, pp. 435–450; también *Mortal Questions*, Cambridge University Press, Cambridge, 1979, pp. 165–180. [Incluido en esta antología, pp. 45–63.]
Nemirow, L., 1989, "Physicalism and the Cognitive Role of Acquaintance", en Lycan 1989, .
———, 1980, Reseña de *Mortal Questions* de T. Nagel, *Philosophical Review*, vol. 89, pp. 473–477.
———, 1979, *Functionalism and the Subjective Quality of Experience*, Stanford (tesis de doctorado).
Tye, M., 1986, "The Subjective Qualities of Experience", *Mind*, vol. 95, pp. 1–17.
Unger, P., 1966, "On Experience and the Development of the Understanding", *American Philosophical Quarterly*, vol. 3, pp. 1–9.

LO QUE NO SON LOS ESTADOS PSICOLÓGICOS*

NED BLOCK Y JERRY FODOR

I

Hasta donde se sabe, organismos diferentes a menudo se hallan en estados psicológicos de exactamente el mismo tipo en uno u otro momento, y un organismo dado se encuentra a menudo en estados psicológicos de exactamente el mismo tipo en momentos diversos. Siempre que se dé uno de estos casos, diremos de los estados psicológicos del organismo o de los organismos en cuestión que son de *tipo idéntico*.

Una de las tareas que frecuentemente intentan realizar las teorías de la filosofía de la mente que hoy en día están en boga es caracterizar las condiciones para la identidad tipo de los estados psicológicos. Por ejemplo, ciertas versiones del conductismo filosófico afirman que dos organismos se encuentran en estados psicológicos de tipo idéntico si y sólo si algunas de sus conductas o disposiciones conductuales son de tipo idéntico. En forma análoga, algunas (aunque no todas) las versiones del fisicalismo aseveran que los organismos se encuentran en estados

* Varios amigos y colegas han leído los primeros borradores. Estamos particularmente en deuda con los profesores Richard Boyd, Donald Davidson, Michael Harnish y Hilary Putnam por la atención con que leyeron el escrito y por sus útiles sugerencias.

psicológicos de tipo idéntico si y sólo si algunos de sus estados físicos son de tipo idéntico.[1]

Mientras se interpreten como teorías sobre las condiciones de identidad tipo de los estados psicológicos, parece cada vez más improbable que el conductismo o el fisicalismo sean verdaderos. Ya que los argumentos a favor de esta conclusión están presentes en una amplia gama de literatura, expondremos aquí sólo una brevísima revisión.[2]

El argumento fundamental contra el conductismo consiste, sencillamente, en que lo que un organismo hace o está dispuesto a hacer en un momento dado es una función muy compleja de sus creencias y deseos, junto con sus entradas sensoriales presentes y sus recuerdos. Es, por lo tanto, extremadamente improbable que resulte posible casar predicados conductuales con predicados psicológicos en la forma exigida por el conductismo; a saber, que para cada tipo de estado psicológico, un organismo se halla en ese estado si y sólo si un predicado conductual específico es verdadero de él. Esto sugiere que es altamente probable que el conductismo sea falso simplemente en virtud de sus consecuencias empíricas e independientemente de su poca plausibilidad como tesis semántica. El conductismo no puede ser verdadero a menos que el correlacionismo entre la mente y la conducta sea verdadero, y el correlacionismo entre la mente y la conducta no es verdadero.

El argumento contra el fisicalismo descansa en la probabilidad empírica de que criaturas de composición y estructu-

[1] Si el fisicalismo es la doctrina de que los estados psicológicos son estados físicos, tenemos entonces dos versiones, dependiendo de si interpretamos a "estados" como refiriéndose a tipos o a casos. La última interpretación ofrece una teoría más débil al suponer que un caso del tipo *x* puede ser idéntico a un caso del tipo *y* a pesar de que *x* y *y* sean tipos distintos. Según este supuesto, el fisicalismo tipo implica claramente el fisicalismo caso, pero no a la inversa.

La distinción entre teorías de identidad caso y las teorías de identidad tipo no ha sido explotada en el caso de los análisis conductuales. Al contrario de ambas versiones del fisicalismo, el conductismo se sostiene por lo general como una tesis semántica, por lo tanto, como una teoría sobre las relaciones lógicas entre tipos. En este artículo, "fisicalismo" significará fisicalismo tipo. Al hablar de los estados especificaremos si nos estamos refiriendo a tipos o casos sólo cuando no resulte claro a partir del contexto.

[2] Véase Davidson 1970; Fodor 1968; Putnam 1965; 1966; 1967.

ra diferentes, que no están en un sentido interesante en estados físicos idénticos, puedan no obstante estar en estados psicológicos idénticos; por ende, de que los tipos de estados psicológicos no guarden correspondencia con los tipos de estados físicos. Este argumento ha sido desarrollado convincentemente en Putnam 1967. En esencia, recurre a las tres siguientes clases de consideraciones empíricas.

En primer lugar, la doctrina de Lashley sobre la equipotencialidad neurológica sostiene que cualquiera de una amplia gama de funciones psicológicas puede ser cumplida mediante cualquiera de una amplia gama de estructuras cerebrales. Si bien la generalidad de esta doctrina es cuestionable, sí parece claro que el sistema nervioso central es sumamente versátil y que un tipo dado de proceso psicológico está frecuentemente asociado con una variedad de estructuras neurológicas distintas. (Por ejemplo, es un hecho ampliamente conocido que un trauma a edad temprana puede conducir al establecimiento de funciones lingüísticas en el hemisferio *derecho* de personas diestras.) Pero el fisicalismo, según lo hemos entendido, requiere que los organismos estén en estados psicológicos de tipo idéntico si y sólo si están en estados físicos de tipo idéntico. Por ende, si la equipotencialidad es verdadera, el fisicalismo tiene que ser falso.

La segunda consideración depende del supuesto de que la doctrina darwiniana de la convergencia se aplica a la filogenia de la psicología, así como a la filogenia de la morfología y del comportamiento. Es bien sabido que similitudes morfológicas superficiales entre los organismos pueden no representar más que soluciones evolutivas paralelas al mismo problema ambiental: en particular, que pueden ser la expresión de tipos completamente distintos de estructuras fisiológicas. La idea análoga acerca de las similitudes conductuales entre especies ha sido ampliamente reconocida en la literatura etológica: organismos de filogenia y morfología muy diferentes pueden llegar, no obstante, a exhibir similitudes conductuales superficiales en respuesta a presiones ambientales convergentes. La idea que nos interesa aquí es que bien pueden aplicarse las mismas consideraciones a la filogenia de la psicología de los organismos. Las similitudes psicológicas entre las especies pueden reflejar a menudo

una selección ambiental convergente en vez de similitudes fisiológicas subyacentes. Por ejemplo, carecemos de una razón especial para suponer que la fisiología del dolor en el hombre deba tener mucho en común con la fisiología del dolor en especies filogenéticamente alejadas de él. Pero si existen organismos cuya psicología es homóloga a la nuestra, pero cuya fisiología es totalmente distinta, tales organismos ofrecen contraejemplos a las correlaciones psicofísicas que requiere el fisicalismo.

Por último, si admitimos la posibilidad conceptual de que los predicados psicológicos puedan aplicarse a los artefactos, parece entonces probable que el fisicalismo resulte empíricamente falso. Pues parece probable que, dada una correlación psicofísica cualquiera que vale para un organismo, es posible fabricar una máquina que sea psicológicamente semejante a ese organismo, pero que sea fisiológicamente lo suficientemente distinta de ese organismo como para que la correlación psicofísica no valga para la máquina.

Lo que estos argumentos parecen mostrar es que las condiciones que el conductismo y el fisicalismo intentan imponer sobre la identidad tipo de los estados psicológicos de los organismos son, en un sentido pertinente, insuficientemente abstractas. Parece probable que organismos que difieren en su comportamiento o en sus disposiciones conductuales puedan hallarse no obstante en estados psicológicos de tipo idéntico, al igual que pueden estarlo organismos en estados físicos distintos. (Discutiremos más adelante un punto de vista "funcionalista" sobre la identidad tipo, el cual intenta fijar los criterios de identidad en un nivel más abstracto del que reconocen el fisicalismo o el conductismo.)

Desde luego, es *posible* que resulte que se dan las correspondencias entre tipos exigidas por el conductismo o por el fisicalismo. El argumento aquí es que, incluso si los estados conductuales o físicos *están* en correspondencia uno a uno con los estados psicológicos, no tenemos hasta ahora ninguna evidencia de que esto sea así; por lo tanto, carecemos de justificación para adoptar teorías filosóficas que lo *exijan*. La paradoja del conductismo y el fisicalismo consiste en que, mientras que la mayor parte de los argumentos que rodean a estas doctrinas

son estrictamente "conceptuales", parece cada vez más probable que las refutaciones decisivas de las mismas sean empíricas.

Suele sugerirse que se podrá responder a estos argumentos suponiendo que, aunque ni los estados conductuales ni los físicos se corresponden uno a uno con los estados psicológicos, puede empero haber una correspondencia de muchos a uno. Esto es, se supone que para cada tipo de estado psicológico existe una disyunción distinta de tipos de estados conductuales (o físicos) tal que un organismo está en ese estado psicológico si y sólo si está en uno de los disyuntos.

No obstante, esta clase de propuesta se topa con serias dificultades. En primer lugar, no es obvio que haya, de hecho, una disyunción *distinta* de estados conductuales (o físicos) que corresponda con cada estado psicológico. Por ejemplo, no hay en realidad ninguna razón para creer que la clase de tipos de conducta que, en la historia total del universo, han expresado (o expresarán) ira en tal o cual organismo sea distinta de la clase de tipos de conductas que han expresado, digamos, dolor. Al considerar esta posibilidad, se debe tener en cuenta que casi cualquier conducta puede convertirse, en las circunstancias adecuadas, en la expresión convencional de prácticamente cualquier estado psicológico, y que un organismo dado que se encuentra en un estado psicológico dado puede exhibir prácticamente cualquier disposición conductual, dependiendo de sus creencias y preferencias. La misma clase de razonamiento sirve, *mutatis mutandis*, para refutar el supuesto de que existe una disyunción distinta de tipos de estados físicos para cada tipo de estado psicológico, pues parece plausible que prácticamente cualquier tipo de estado físico pueda realizar prácticamente cualquier tipo de estado psicológico en una u otra clase de sistema físico.

Pero incluso si *hay* realmente una disyunción distinta de tipos de estados conductuales (o físicos) que corresponda con cada tipo de estado psicológico, no habría ninguna razón para creer que tal correspondencia es legaliforme; y no queda claro qué interés filosófico habría en el descubrimiento de una propiedad conductual (o física) que accidentalmente resultara ser coextensiva con un predicado psicológico. Así, como Davidson ha señalado, según el supuesto de que las correlaciones psicoconductuales no son legaliformes, aunque "encontráramos una ora-

ción abierta formulada en términos conductuales y exactamente coextensiva con un predicado mental, no podría razonablemente convencernos de que la hemos encontrado" (Davidson 1970). Y como también lo señaló Davidson, la misma observación se aplica, *mutatis mutandis*, a los predicados físicos.

Finalmente, una teoría que diga que cada predicado psicológico es coextensivo con una disyunción distinta de predicados conductuales (o físicos)[3] es incompatible con lo que hemos supuesto que es una verdad evidente, a saber, que un determinado estado conductual puede expresar (o un determinado estado físico puede realizar) diferentes estados psicológicos en diferentes momentos. Supongamos, por ejemplo, que tenemos una teoría que establece que el predicado psicológico p_1 es coextensivo con el predicado conductual disyuntivo α y que el predicado psicológico p_2 es coextensivo con el predicado conductual disyuntivo β. Supongamos, además, que S_i designa un tipo de conducta que en algunas ocasiones ha expresado p_1 pero no p_2, y que en otras ha expresado p_2 pero no p_1. Así pues, S_i tendrá que ser un disyunto tanto de α como de β. Pero los disyuntos de α son cada uno condiciones suficientes de p_1 y los disyuntos de β son cada uno condiciones suficientes de p_2, bajo el supuesto de que p_1 y α, y p_2 y β, son, respectivamente, coextensivos. Luego, la teoría implica que un organismo en S_i está tanto en p_1 como en p_2, lo cual es lógicamente incompatible con la afirmación de que S_i expresa p_1 (pero no p_2) en ciertas ocasiones y p_2 (pero no p_1) en otras. Desde luego, se podría eludir esta objeción incluyendo designadores espaciotemporales en la especificación de los disyuntos mencionados en α y β. Pero hacer esto sería abandonar por completo el proyecto de expresar correlaciones psicoconductuales (o psicofísicas) mediante bicondicionales legaliformes.

[3] No todos los conductistas filosóficos mantienen esta posición; el conductismo filosófico puede ser caracterizado en términos generales como la posición de que para cada predicado psicológico existe un predicado conductual con el que guarda una "relación lógica". (Véase Fodor 1968.) Luego, la siguiente idea califica como conductista: todas las adscripciones de predicados psicológicos implican adscripciones de predicados conductuales, pero no a la inversa. Aunque esta forma de conductismo no es vulnerable al presente argumento, los anteriores la rebaten con tanta eficacia como a las formas bicondicionales del conductismo.

II

Se ha propuesto recientemente que es posible sortear esta clase de dificultades y formular una teoría adecuada acerca de las condiciones para la identidad tipo de los estados psicológicos de la siguiente manera: supongamos que cualquier sistema P, al cual pueden aplicarse predicados psicológicos, puede describirse como un autómata probabilístico. (Un autómata probabilístico es una máquina de Turing generalizada, cuya tabla de máquina contiene instrucciones asociadas con probabilidades positivas finitas que son menores que o iguales a uno. Para una breve introducción al concepto de una máquina de Turing, una tabla de máquina y otras nociones afines, véase Putnam 1967.) Una *descripción* de P, en el sentido técnico que aquí pretende dársele, es cualquier enunciado verdadero de que P posee distintos estados $S_1, S_2 \ldots S_n$, los cuales están relacionados entre sí y con las salidas y entradas de P por medio de las probabilidades de transición dadas en una tabla de máquina específica. Llamaremos a los estados $S_1, S_2 \ldots S_n$, especificados por la *descripción* de un organismo, los "estados de la tabla de máquina del organismo" correspondientes a esa *descripción*.

Es en el contexto del supuesto de que los organismos pueden ser descritos como autómatas, que esta teoría (en adelante "TIEF", por "teoría de identidad de los estados funcionales") busca especificar las condiciones para la identidad tipo de los estados psicológicos. En particular, la TIEF afirma que, para cualquier organismo que satisfaga predicados psicológicos, existe una *descripción* única óptima, conforme a la cual cada estado psicológico del organismo es idéntico a uno de sus estados de tabla de máquina relativos a esa descripción.

Son pertinentes varias observaciones acerca de la TIEF. En primer lugar, hay una generalización obvia de la noción de autómata probabilístico, en la cual se considera que tiene una cinta de entradas independiente sobre la que un "oráculo" puede imprimir símbolos durante un proceso computacional. La TIEF presupone una interpretación de esta generalización, según la cual unos transductores sensoriales toman el lugar del "oráculo" y las salidas se consideran como instrucciones a los transductores motrices. Una interpretación semejante debe esperarse si

una *descripción* de un organismo ha de proporcionar un modelo de las operaciones mentales del organismo.

En segundo lugar, hemos presentado la TIEF de la manera habitual, como una teoría de *identidad*:[4] en particular, como una teoría que afirma que cada tipo de estado psicológico es idéntico a (un tipo de) estado de tabla de máquina. Sin embargo, nuestro objetivo es dejar de lado las interrogantes acerca de las condiciones de identidad de los objetos abstractos y discutir sólo cierta clase de bicondicionales que los enunciados de identidad tipo-tipo implican: esto es, enunciados de la forma "*O* presenta tal y cual tipo de estado psicológico en el tiempo *t* si y sólo si *O* está en tal y cual tipo de estado de tabla de máquina en el tiempo *t*".

En tercer término, vale la pena insistir en que la TIEF es más que la afirmación de que todo organismo posee una descripción como máquina de Turing o como autómata probabilístico. Pues existen varios sentidos en los cuales esa afirmación es trivialmente verdadera; pero su verdad en estos sentidos no implica la TIEF. Por ejemplo, si las entradas y salidas de un organismo son recursivamente enumerables (como sucede con cualquier organismo mortal después de su fallecimiento), se sigue entonces que existe una máquina de Turing capaz de simular al organismo (es decir, una máquina de Turing que tiene las mismas relaciones entre entradas y salidas). Pero no se sigue que el organismo posee una *descripción* única óptima del tipo caracterizado más arriba. En segundo lugar, como lo ha señalado Putnam (en conversación), *todo* es describible como una realización de lo que podría llamarse máquina de Turing "nula", esto es, una máquina que sólo posee un estado y permanece en él. (El argumento consiste, toscamente, en que el que un sistema *P* realice o no una máquina de Turing depende, entre otras cosas, de lo que cuenta como un cambio de estado en *P*. Si se hace valer *nada* como un cambio de estado en *P*, entonces *P* constituye una realización de la máquina de Turing nula.) Pero nuevamente, la TIEF no sería verdadera si la única *descripción* verdadera de un organismo fuera una máquina de Turing nula, ya que la TIEF requiere que los estados de tabla de máquina de un organismo

[4] *Cfr.* Putnam 1967 y 1970.

correspondan uno a uno con sus estados psicológicos bajo su descripción óptima.

Hay así dos sentidos importantes en los cuales la TIEF involucra más que la afirmación de que los organismos que satisfacen predicados psicológicos poseen descripciones. En primer lugar, la TIEF afirma que tales sistemas poseen descripciones únicas óptimas. En segundo lugar, la TIEF afirma que los tipos de estados de tabla de máquina especificados por la descripción única óptima de un sistema corresponden a los tipos de estados psicológicos en que puede estar el sistema. Es esta segunda afirmación de la TIEF la que constituirá nuestro principal interés.

La TIEF, a diferencia del conductismo o del fisicalismo, no es una teoría ontológica: esto es, es neutral con respecto a lo que *son* los estados psicológicos caso, por cuanto, hasta donde le interesa a la TIEF, entre los sistemas a los cuales se aplican los predicados psicológicos (y que en consecuencia poseen *descripciones*) pueden incluirse personas, objetos materiales, almas y otros. Esta última idea indica cómo la TIEF es capaz de soslayar algunos de los tipos de dificultades que planteamos con respecto al fisicalismo y el conductismo. Así como la TIEF se abstrae de las consideraciones acerca del estatus ontológico de los sistemas que poseen *descripciones*, también se abstrae de las diferencias físicas entre sistemas que poseen *descripciones* comunes. Como lo ha observado Putnam, "la *misma* máquina de Turing (desde el punto de vista de la tabla de máquina) puede realizarse físicamente en una variedad potencialmente infinita de formas" (1966, p. 271), y la TIEF nos permite enunciar condiciones para la identidad tipo de estados psicológicos que son neutrales respecto de esas realizaciones diferentes.

Asimismo, la TIEF nos permite enunciar dichas condiciones en términos que se abstraen de la variedad de consecuencias conductuales que pueda tener un estado psicológico. De este modo, supera un tipo de objeción que, como supusimos más arriba, era fatal para el conductismo.

Observamos que el conductista está comprometido con la idea de que dos organismos se encuentran en el mismo estado psicológico siempre que sus conductas y/o disposiciones conductuales sean idénticas; y que esta teoría es poco plausible en

la medida en que las conductas y las disposiciones conductuales de un organismo son efectos de las *interacciones* entre sus estados psicológicos. Pero la TIEF nos permite distinguir entre los estados psicológicos no sólo en términos de sus consecuencias conductuales, sino también en términos del carácter de sus interconexiones. Esto se debe a que el criterio de identidad para los estados de tabla de máquina reconoce *las relaciones entre ellos*, además de sus relaciones con las entradas y salidas. Así, la TIEF puede hacer frente al carácter indirecto distintivo de la relación entre estados psicológicos y comportamiento. De hecho, la TIEF nos permite ver cómo estados psicológicos que *no* tienen una expresión conductual pueden ser, no obstante, distintos.

Finalmente, cabe notar que nada impide tomar por lo menos algunas de las transiciones especificadas en una tabla de máquina como si correspondiera a las relaciones causales del sistema que la tabla *describe*. En particular, puesto que la TIEF es compatible con el fisicalismo caso, no hay motivos para que no reconozca que los estados psicológicos caso pueden entrar en relaciones causales. Así, cualquier ventaja que redunde en beneficio de los análisis causales de los estados psicológicos, o de las relaciones entre estados psicológicos y conducta, redundará igualmente en beneficio de la TIEF.[5]

III

En esta sección vamos a revisar una serie de argumentos que ofrecen cierto grado de dificultad a la afirmación de que la TIEF proporciona una explicación adecuada de las condiciones para la identidad tipo de los estados psicológicos. Nuestro punto de vista es que, tomados en conjunto, estos argumentos son suficientemente decisivos contra la teoría de identidad tipo de los estados psicológicos propuesta por la TIEF. En la sección final indicaremos algunas razones acerca de por qué el intento por ofrecer condiciones sustanciales para la identidad tipo de los estados psicológicos fracasa con tanta frecuencia.

(1) Cualquier explicación de las condiciones para la identidad tipo de los estados psicológicos que pretenda adherirse

[5] *Cfr.* Davidson 1963.

estrechamente a nuestra noción cotidiana de qué tipos de estados psicológicos hay seguramente tendrá que establecer una distinción entre estados disposicionales (creencias, deseos, inclinaciones, etc.) y estados incidentes* (sensaciones, pensamientos, emociones, etc.). Sin embargo, hasta donde se alcanza a ver, la TIEF carece de una vía plausible para captar esta distinción a menos que abandone su principio fundamental de que los estados psicológicos se corresponden uno a uno con los estados de la tabla de máquina. Supongamos, por ejemplo, que la TIEF intenta reconstruir la distinción entre incidentes y disposiciones haciendo referencia a la distinción entre el estado de la tabla de máquina *en* el que está un organismo y todos los otros estados especificados por su tabla de máquina. Así, se podría pulir la TIEF de suerte que dijera: para los estados incidentes, dos organismos se encuentran en estados psicológicos de tipo idéntico si y sólo si están en el mismo estado de tabla de máquina; y para cada estado disposicional, existe un estado de tabla de máquina tal que un organismo está en el primero si y sólo si su tabla de máquina contiene el segundo.

Lo siguiente indica una forma de poner de manifiesto la poca plausibilidad de esta propuesta. Todo estado de tabla de máquina de un organismo es un estado en el que ese organismo puede estar en uno u otro momento. Por lo tanto, si la distinción entre el estado de tabla de máquina en el que está un organismo y todos los otros estados de su tabla de máquina es igual a la distinción entre el estado incidente de un organismo y sus estados disposicionales, se sigue que todo estado disposicional del organismo es un estado incidente posible de ese organismo.

Esta consecuencia de la TIEF es correcta para numerosas clases de disposiciones psicológicas. Por ejemplo, en correspondencia con el predicado disposicional "habla francés", tenemos el predicado incidente "está hablando francés"; en correspondencia con el predicado disposicional "es voraz", tenemos el predicado incidente "está siendo voraz"; en correspondencia con el predicado disposicional "puede oír sonidos mayores de 3 000 hercios", tenemos el predicado incidente "está oyendo un

* Hemos traducido "*ocurrent*" como "incidente" tanto en su forma adjetival como sustantiva. [N. de las comps.]

sonido mayor a 3 000 hercios". Y, en general, para muchos predicados disposicionales, tenemos las formas correspondientes en tiempo presente continuo que denotan incidencias.

Sin embargo, para muchos otros predicados disposicionales psicológicos, este paralelismo no se sostiene. Por ejemplo, carecemos de un predicado "está creyendo que *P*" que corresponda a "cree que *P*"; carecemos de "está deseando un caramelo" que corresponda con "desea un caramelo"; carecemos de "está prefiriendo *X* sobre *Y*" que corresponda con "prefiere *X* a *Y*", y así sucesivamente. En suma, muchos estados psicológicos disposicionales *no* son estados psicológicos incidentes posibles, y para éstos la TIEF no ofrece un modelo claro.

Es importante entender lo que este argumento *no* demuestra. De acuerdo con este argumento, ciertos estados disposicionales no pueden corresponder a estados de tabla de máquina porque todos los estados de tabla de máquina son estados incidentes posibles, pero algunos estados psicológicos disposicionales no lo son. Para esas disposiciones, no puede haber estados en la tabla de máquina que permitan afirmar que el organismo tiene el primero si y sólo si aparece el segundo en su *descripción*. Sin embargo, es perfectamente posible que las condiciones necesarias y suficientes para tener tales disposiciones deban darse por referencia a alguna propiedad *abstracta* de la organización de las tablas de máquina. Tomando un ejemplo improbable, dada una forma normal de las descripciones, podría resultar que un organismo creyera que el Sol se encuentra a 149.5 millones de kilómetros de la Tierra si y sólo si las primeras *n* columnas de su tabla de máquina poseen una propiedad abstracta tal que contiene únicamente números enteros nones. Puesto que decir de una máquina que las primeras *n* columnas..., etc., no es adscribirle un estado de tabla de máquina, los estados psicológicos que se analizan como correspondientes a esta clase de propiedad no se describirían por ello como estados incidentes posibles.

Sin embargo, seguir este camino implicaría abandonar una afirmación fundamental de la TIEF. Pues, aunque esta manera de abordar el asunto resulta compatible con la idea de que dos organismos tienen la misma psicología si y sólo si tienen la misma tabla de máquina, no es compatible con la sugerencia

de que dos organismos se encuentran en el mismo estado psicológico (disposicional) si y sólo si tienen en común un estado especificado en sus tablas de máquina. Por ende, es incompatible con la idea de que los estados psicológicos corresponden uno a uno con los estados de tabla de máquina. Además, como no tenemos manera de saber qué clases de propiedades abstractas de las tablas de máquina podrían llegar a corresponder con los estados psicológicos, la dificultad presente reduce en gran medida la posibilidad de usar la TIEF para delinear condiciones sustanciales para la identidad tipo de los estados psicológicos. Decir que los estados psicológicos corresponden a una u otra propiedad de las tablas de máquina es decir algo mucho más débil que decir que los estados psicológicos corresponden a estados de la tabla de máquina. Éste es un tipo de argumento al que volveremos más adelante en la discusión.

Hay, desde luego, por lo menos otro camino para salir de la presente dificultad de la TIEF. Podría sugerirse que debemos abandonar la noción cotidiana de que hay ciertos estados disposicionales que no son estados incidentes posibles (por ejemplo, reconocer un estado incidente, aunque tal vez no consciente, de estar creyendo que *P*). Está claro que la posibilidad de que un día podríamos tener bases teóricas para reconocer la existencia de tales estados no puede excluirse *a priori*. Pero carecemos de tales bases *en este momento*, y nos parece que sí hay un principio metodológico del conservacionismo que establece que debemos evitar los modelos que exigen cambios empíricos o conceptuales que no estén motivados de manera independiente.

(2) Sugerimos que la TIEF nos permite explicar el hecho de que la conducta es típicamente el producto de las interacciones entre estados psicológicos, y que la existencia de tales interacciones constituye una fuente constante de dificultades para las teorías conductistas, en cuanto éstas buscan asignar conductas características *individualmente* a estados psicológicos. Desde el punto de vista empírico, sin embargo, es sumamente probable que haya *dos* clases de interacciones conductualmente eficaces entre los estados psicológicos, y la TIEF sólo ofrece un modelo natural para una de ellas.

Por un lado, el comportamiento puede ser producto de una *serie* de estados psicológicos, y la explicación de la TIEF nos mues-

tra de qué manera podría resultar esto cierto y cómo algunos de los estados que ocurren en esa serie pueden ellos mismos carecer de expresiones conductuales. Pero, por otro lado, la conducta puede ser el resultado de interacciones entre estados mentales *simultáneos*. Por ejemplo, a primera vista, lo que un organismo hace en *t* puede ser una función de lo que siente en *t* y de lo que piensa en *t*. Pero la TIEF no proporciona ninguna herramienta conceptual que nos permita representar este estado de cosas. En efecto, la TIEF es capaz de ofrecer la representación de interacciones secuenciales entre estados psicológicos, pero no de interacciones simultáneas. De hecho, la TIEF ni siquiera da cuenta del hecho de que un organismo puede estar en más de un estado psicológico incidente al mismo tiempo, ya que un autómata probabilístico sólo puede estar en un estado de tabla de máquina a la vez. La consecuencia de este argumento parece ser que si ha de utilizarse a los autómatas probabilísticos como modelos de un organismo, el modelo adecuado es un conjunto de autómatas intercomunicados que operen en forma paralela.

También es importante tener en claro lo que este argumento no demuestra con respecto a la TIEF. Hemos interpretado la TIEF como si afirmara que los estados psicológicos de un organismo guardan una correspondencia uno a uno con los estados de tabla de máquina postulados en su *descripción* óptima. El presente argumento indica que, para aceptar esta afirmación, la *descripción* óptima de un organismo no debe representarlo como un autómata probabilístico individual. Si los organismos, pero no los autómatas probabilísticos individuales, pueden estar en más de un estado a la vez, entonces o bien un organismo no es un autómata probabilístico individual, o bien los estados psicológicos de un organismo no corresponden a los estados de tabla de máquina de los autómatas probabilísticos individuales. (Debe señalarse que existe un algoritmo para construir un autómata individual que sea equivalente a cualquier conjunto dado de autómatas paralelos. Sin embargo, no puede darse el caso de que *tanto* un conjunto de autómatas paralelos *como* el autómata individual equivalente ofrezcan a la vez *descripciones* óptimas de un organismo.)

Estas observaciones poseen cierta importancia, ya que el tipo de teoría psicológica que obtenemos a partir del supuesto de

que los organismos son procesadores paralelos resulta totalmente distinta del tipo que obtenemos a partir del supuesto de que son procesadores en serie. De hecho, aunque las características generales de los procesadores en serie se comprenden relativamente bien, se sabe muy poco de las correspondientes características de los sistemas paralelos.

Por otra parte, este argumento deja incólume la afirmación principal de la TIEF: incluso si los organismos son, en cierto sentido, conjuntos de autómatas probabilísticos, puede ocurrir que cada uno de los estados psicológicos de un organismo corresponda al estado de tabla de máquina de uno u otro miembro del conjunto. En los argumentos que siguen, asumiremos el modelo en serie por razones de simplicidad e intentaremos mostrar que, incluso bajo ese supuesto, los estados psicológicos no corresponden a los estados de la tabla de máquina.

(3) La TIEF sostiene que dos organismos están en estados psicológicos del mismo tipo si y sólo si están en el mismo estado de tabla de máquina. Pero los estados de tabla de máquina son idénticos si y sólo si están relacionados de modo idéntico con otros estados de tabla de máquina y con los estados de los mecanismos de entrada y salida. En este sentido, el criterio para la identidad de los estados de tabla de máquina es la "equivalencia funcional". Por lo tanto, la TIEF afirma que la identidad tipo de los estados psicológicos es igualmente una cuestión de cierta clase de equivalencia funcional; los estados psicológicos son de tipo idéntico si y sólo si comparten las propiedades que deben especificarse para individuar un estado de tabla de máquina.

Sin embargo, podría argumentarse de manera plausible que este modo de identificar en tipos a los estados psicológicos no logra dar cabida a un rasgo de por lo menos algunos de esos estados, el cual es crucial para determinar su tipo: a saber, su carácter "cualitativo". No parece, por ejemplo, totalmente irrazonable sugerir que no habría ningún caso del tipo "estado de dolor" a menos que se sintiera como dolor, y que esto sería verdad aun cuando estuviera relacionado con todos los otros estados psicológicos del organismo de cualesquiera que sean las formas en que los dolores están relacionados. Nos parece que los contraargumentos verificacionistas estándares que se oponen a la idea de que la hipótesis del "espectro invertido" es

conceptualmente incoherente son poco convincentes. Si esto es correcto, daría entonces la impresión de que la posibilidad de la inversión de los *qualia* representa un argumento *prima facie* serio contra las explicaciones funcionalistas de los criterios para la identidad tipo de los estados psicológicos.

Debe observarse, sin embargo, que el argumento de la inversión de los *qualia* constituye *sólo* una objeción *prima facie* contra la TIEF. En particular, el proponente de las explicaciones funcionalistas dispone de cualquiera de dos medios para hacer frente a esta objeción. Por un lado, podría argumentar que, aunque los *qualia* invertidos, *si se dieran*, ofrecerían contraejemplos de su teoría, es imposible, nomológicamente, que estados psicológicos funcionalmente idénticos sean cualitativamente distintos: en particular, que cualquier cosa que alterase las características cualitativas de un estado psicológico alteraría sus características funcionales. Esta línea de razonamiento podría parecer un flagrante apriorismo, pero, en ausencia de los datos empíricos pertinentes, bien puede adoptarse la actitud de esperar y ver.

Hay, además, otra defensa posible para el proponente de la TIEF. Puede decir que, dados dos estados psicológicos funcionalmente idénticos, podríamos (o quizá "deberíamos") *tomarlos* como idénticos en tipo, independientemente de sus propiedades cualitativas: es decir, que las diferencias entre las propiedades cualitativas de los estados psicológicos que no determinen diferencias funcionales correspondientes carecen *ipso facto* de pertinencia para los objetivos de la construcción de teorías en psicología y, por lo tanto, deben ignorarse al establecer identificaciones tipo.

Para apreciar la plausibilidad de esta sugerencia, imaginemos que se descubriera que cada persona tiene, de hecho, *qualia* ligeramente distintos (o, mejor aún, *qualia* extremadamente distintos) cuando está en cualquier estado de tabla de máquina que se supone que es idéntico al dolor. Parece bastante claro que, en este caso, podría ser razonable afirmar que el carácter de los *qualia* de un organismo carece de pertinencia con respecto a si tiene un dolor o (en términos equivalentes) que los dolores se sienten de manera muy distinta en organismos diferentes.

Esta forma de argumento puede conducir, no obstante, a consecuencias vergonzosas. Hasta donde sabemos ahora, puede ser nomológicamente posible que dos estados psicológicos sean funcionalmente idénticos (es decir, que estén relacionados idénticamente con las entradas, las salidas y los estados sucesores), aun cuando sólo uno de los estados posea un contenido cualitativo. En este caso, la TIEF nos obligaría a decir que un organismo puede tener un dolor aunque no esté sintiendo *absolutamente nada*, y esta consecuencia parece totalmente inaceptable.

Cabe observar que estos casos de "*qualia* invertidos (o ausentes)" le plantean, en cierto sentido, un problema más profundo a la TIEF del que le plantean cualquiera de los otros argumentos que discutiremos. Nuestros otros argumentos apuntan, en lo principal, a demostrar que los estados psicológicos no pueden definirse funcionalmente de cierta manera, a saber, estableciendo una correspondencia entre ellos y los estados de tabla de máquina. Pero aunque sean incompatibles con la TIEF, son compatibles con el funcionalismo en el sentido amplio de esa doctrina, la cual sostiene que las condiciones para la identidad tipo de los estados psicológicos refieren sólo a sus relaciones con las entradas, las salidas, y a las relaciones entre sí. Sin embargo, podría tomarse esta consideración como si demostrara que los estados psicológicos no pueden definirse *en lo absoluto* en términos funcionalistas, y que no pueden corresponder a *ninguna* propiedad definible en términos de autómatas abstractos. Haremos caso omiso de esta posibilidad en lo que sigue, ya que si los estados psicológicos no son estados funcionales en lo absoluto, sencillamente no surge la pregunta de si son o no estados de tabla de máquina.

(4) Observamos que hay ciertas objeciones al conductismo y al fisicalismo que indican que cada uno de éstos propone restricciones a las condiciones para la identidad tipo de los estados psicológicos que, de una u otra forma, son insuficientemente abstractas. Argumentaremos ahora que la misma clase de objeción se aplica a la TIEF.

Una tabla de máquina especifica un estado en términos de un conjunto de instrucciones que controlan el comportamiento de la máquina siempre que se encuentre en ese estado. Por

definición, en el caso de un autómata determinista, sus instrucciones especifican, para cada estado de la máquina, una salida asociada y un estado de máquina sucesor. El autómata probabilístico difiere del anterior sólo en que cualquier estado puede especificar una *gama* de salidas o de estados sucesores, con una distribución de probabilidades asociada. En suma, dos estados de tabla de máquina de un autómata determinista son distintos si difieren en sus salidas asociadas o en su estado sucesor asociado. De manera análoga, dos estados de tabla de máquina de un autómata probabilístico difieren, si difieren en su gama de salidas o en su gama de estados sucesores o en las distribuciones de probabilidades asociadas con alguna de estas gamas.

Si, no obstante, trasladamos esta convención para distinguir estados de tabla de máquina a la identificación tipo de los estados psicológicos, obtenemos condiciones de identidad que son, por así decirlo, demasiado finas. De este modo, por ejemplo, si tú y yo *sólo* diferimos en que tu respuesta más probable al dolor causado por pegarse en el dedo del pie es exclamar "maldición", mientras que la mía es decir "maldita sea", se sigue que el dolor que sientes al pegarte en el dedo del pie es de tipo distinto del que yo siento cuando me pego en el dedo del pie.

Este argumento se repite de manera vergonzosa. Para ver esto, consideremos el caso especial del autómata determinista: x y y son estados de tabla de máquina de distinto tipo de tal autómata si los estados sucesores inmediatos de x y y son de tipo distinto. Pero los estados sucesores inmediatos de x y y son de tipo distinto si *sus* estados sucesores inmediatos son de tipo distinto. De este modo, x y y son de tipo distinto si los sucesores inmediatos de sus sucesores inmediatos son de tipo distinto, y así sucesivamente. De hecho, bajo el supuesto de que existe una vía computacional que lleva de cada estado a todos los demás, dos autómatas cualesquiera que tengan menos que todos sus estados en común no tendrán ningún estado en común. Este argumento se generaliza a los autómatas probabilísticos de manera obvia.

Nuevamente, es importante observar lo que *no* demuestra este argumento. En particular, no demuestra que los estados psicológicos no puedan identificarse a nivel de tipos haciendo referencia a cierta clase de propiedades *abstractas* de los estados

de tabla de máquina. Pero, como señalamos en nuestra discusión sobre el argumento 1, decir que los estados psicológicos corresponden a tal o cual propiedad definible en los estados de tabla de máquina es decir mucho menos acerca de las condiciones para la identidad tipo de los estados psicológicos que lo que pretende decir la TIEF. Y el argumento presente sí parece mostrar que las condiciones utilizadas para la identificación tipo de los estados de tabla de máquina *per se* no pueden emplearse para la identificación tipo de los estados psicológicos. Es de suponerse que este tipo de razonamiento es el que tenía en mente Putnam, por ejemplo, cuando advierte que "la dificultad consistirá, desde luego, en pasar de los modelos de los organismos *específicos* a una forma *normal* de la descripción psicológica de los *organismos*" (Putnam 1967, p. 43). En suma, a primera vista puede parecer que la explotación de los criterios empleados para la identificación tipo de los estados de tabla de máquina ofrece a la TIEF conceptos que se encuentran precisamente en el nivel de abstracción requerido para la identificación tipo de los estados psicológicos. Pero, de hecho, esto no parece ser cierto.

(5) Nos parece que el siguiente argumento muestra que no es posible establecer una correspondencia uno a uno entre los estados psicológicos de los organismos y los estados de tabla de máquina de los organismos.

El conjunto de estados que constituyen la tabla de máquina de un autómata probabilístico es, por definición, una lista. Pero el conjunto de estados mentales de, por lo menos, algunos organismos (a saber, de personas) es, como una cuestión de hechos empíricos, productivo. En particular, haciendo abstracción de las limitaciones impuestas por la memoria y la mortalidad, que no son teóricamente pertinentes, hay una variedad infinita de estados psicológicos de cualquier persona dada que son de distinto tipo y que son nomológicamente posibles. La demostración más sencilla de que esto es cierto es que, bajo el supuesto de que hay una variedad infinita de oraciones declarativas que no son equivalentes, se pueden generar descripciones definidas de esos estados sustituyendo la *S* por oraciones en el esquema *A*:

A: "la creencia (pensamiento, deseo, esperanza, o lo que fuere) de que *S*"

En suma, mientras que el conjunto de estados de tabla de una máquina de Turing puede, por definición, especificarse exhaustivamente mediante una lista, el conjunto de estados mentales de una persona puede, en el mejor de los casos, especificarse mediante una axiomatización finita.

A este argumento se podría objetar que no más de un subconjunto finito de las descripciones definidas generables a partir de la sustitución realizada en *A* designa de hecho las creencias (deseos, esperanzas o lo que fuere) nomológicamente posibles, y que esto es cierto *no* debido a las limitaciones, carentes de interés teórico, impuestas por la memoria y la mortalidad, sino más bien debido a la existencia de leyes psicológicas que limitan el conjunto de proposiciones creíbles (deseables, esperables, o lo que fuere) a un conjunto finito. Tomando un ejemplo improbable, se podría sostener que si se eliminan todas las proposiciones que quizá no son creíbles, como "2 + 2 = 17", "2 + 2 = 147" y demás, el residuo es un conjunto finito.

Sin embargo, no existe razón alguna para suponer que esto sea verdad, y hay razones muy convincentes para creer que no lo es. Por ejemplo, el conjunto infinito de descripciones cuyos miembros son "la creencia de que 1 + 1 = 2", "la creencia de que 2 + 2 = 4", "la creencia de que 3 + 3 = 6", y demás, parecería designar un conjunto de creencias posibles de un organismo que es idealmente libre de limitaciones de memoria; para expresarlo de otro modo, el hecho de que haya enunciados aritméticos que a una persona le son nomológicamente imposibles de creer es consecuencia del carácter de la memoria de la gente, no del carácter de su representación mental de la aritmética.

Debe subrayarse nuevamente que esto pretende ser una afirmación empírica, aunque en extremo plausible. Es posible imaginar a una criatura idealmente libre de limitaciones de memoria cuya representación mental de la aritmética especifica, no obstante, sólo un conjunto finito de creencias aritméticas posibles. El argumento aquí es que resulta sumamente improbable que nosotros seamos tales criaturas.

Una vez más es importante observar lo que este argumento *no* demuestra. Hagamos una distinción entre los *estados de tabla de máquina* de un autómata y los *estados computacionales* de un autómata. Por lo primero, entenderemos lo que hemos venido entendiendo desde el principio: estados especificados por columnas en la tabla de la máquina. Por lo segundo, entenderemos cualquier estado de la máquina que pueda caracterizarse en términos de sus entradas, salidas y/o estados de tabla de máquina. Según esta acepción, los predicados "acaba de realizar un cómputo que entraña trescientos setenta y dos estados de tabla de máquina", o "acaba de demostrar el último teorema de Fermat", o "acaba de teclear el símbolo número 'iésimo' en su vocabulario de salida" designan todos estados computacionales posibles de las máquinas.

Ahora bien, lo que este argumento parece mostrar es que no es posible establecer una correspondencia entre los estados psicológicos de un organismo y los estados de tabla de máquina de un autómata. Lo que, desde luego, *no* demuestra es que no es posible establecer una correspondencia entre los estados psicológicos de un organismo y los estados *computacionales* de un autómata. De hecho, una condición suficiente para la existencia de esta última correspondencia es que los estados psicológicos de un organismo deban ser enumerables.[6]

[6] La afirmación de que los organismos son autómatas probabilísticos puede ser verdadera de manera interesante aunque la TIEF sea falsa; esto es, aun si los estados psicológicos no corresponden a los estados de la tabla de máquina. Por ejemplo, podría suceder que cierto subconjunto de estados psicológicos de un organismo correspondiera a un conjunto de estados de la tabla de máquina mediante los cuales se determinara el resto de su psicología. O podría resultar que lo que corresponde a cada estado de la tabla de máquina sea una conjunción de estados psicológicos..., etc. De hecho, aunque la afirmación de que cualquier organismo puede ser modelado por un autómata probabilístico no sea interesante, la afirmación de que para cada organismo hay un autómata probabilístico que constituye su modelo único óptimo es interesante. Y esta última afirmación no implica la TIEF ni es, en modo alguno, obviamente verdadera.

En suma, hay muchas maneras en que podría resultar que los organismos fueran autómatas en un sentido más interesante que en el sentido en que cada cosa es un autómata conforme a cierta descripción. Nuestro argumento aquí consiste en que tales eventualidades, aunque serían importantes, no ofrecerían

(6) Hemos argumentado que, como el conjunto de estados de tabla de máquina de un autómata no es productivo, no se puede establecer una correspondencia entre ellos y los estados psicológicos de un organismo. Argumentaremos a continuación que, incluso si pudiera llevarse a cabo tal correspondencia, ésta necesariamente fracasaría en representar propiedades esenciales de los estados psicológicos. Parece bastante claro que existen similitudes estructurales por lo menos entre algunos estados psicológicos, y que una teoría exitosa de tales estados debe representar y explotar esas similitudes. Por ejemplo, está claro que hay una relación teóricamente importante entre el estado psicológico de alguien que cree que P y el estado psicológico de alguien que cree que $P \& Q$. El argumento aquí dice simplemente que representar los estados psicológicos como una lista (por ejemplo, como una lista de estados de tabla de máquina) no representa esta clase de relación estructural. Es preciso decir que creer que P es de alguna manera[7] un componente de creer que $P \& Q$; pero el modelo de los estados de tabla de máquina carece de recursos conceptuales para decir esto. En particular, la noción de "es un componente de" no está definida para los estados de tabla de máquina.

Podría replicarse que esta clase de argumento no es estrictamente pertinente en las afirmaciones de la TIEF, pues sin duda es posible, en principio, que haya una correspondencia uno a uno entre los estados de tabla de máquina y los estados psicológicos, aunque el vocabulario adecuado para la individuación de los primeros no capte las relaciones estructurales entre los segundos.

A esta réplica, no obstante, se le escapa el meollo del problema. A fin de entender esto, consideremos el asunto mediante oraciones. La razón de que haya paralelismos estructurales entre las oraciones es que éstas se construyen a partir de un

las condiciones generales para la identificación tipo de estados psicológicos de la manera en que la TIEF pretende hacerlo.

[7] Ciertamente "de alguna manera". Es obvio que creer que p no es un componente de creer que $p \vee q$ del mismo modo en que creer que p es un componente de creer que $p \& q$. Igualmente obvio es que existe cierta relación entre creer que p y creer que $p \vee q$, y una teoría de las creencias tendrá que determinar cuál es esa relación.

conjunto fijo de elementos de un vocabulario por medio de la aplicación iterada de un conjunto fijo de reglas, y el vocabulario teórico requerido para analizar los modos en que las oraciones son estructuralmente semejantes es precisamente el vocabulario requerido para especificar el dominio de esas reglas. En particular, las oraciones que son estructuralmente similares comparten el léxico o las formas de derivación, o ambos. Así, las similitudes estructurales entre oraciones se explican de la misma manera en que se explica su productividad, a saber, describiéndolas como un conjunto generado en vez de a una lista.

Nuestro argumento es que las mismas consideraciones se aplican al conjunto de estados psicológicos de un organismo. Casi seguramente, también éstos son, o por lo menos incluyen, un conjunto generado, y sus similitudes estructurales corresponden, al menos en parte, a similitudes en su derivación; es decir, con los estados psicológicos, al igual que con las oraciones, el hecho de que sean productivos y el hecho de que exhiban una estructura interna son dos aspectos del mismo fenómeno. Si esto es cierto, entonces una teoría que no capte las relaciones estructurales que se dan dentro de los estados psicológicos y entre ellos no tiene prácticamente ninguna posibilidad de llegar a una *descripción* adecuada para los propósitos de la psicología teórica.

Así, este argumento, al igual que el argumento 5, conduce a la conclusión de que, si queremos pensar en la psicología de los organismos como representada por autómatas, entonces los estados psicológicos de los organismos parecen ser análogos a los estados computacionales de un autómata, y no a los estados de tabla de máquina.

IV

Hemos estado considerando teorías de la filosofía de la mente que pueden interpretarse como intentos por establecer condiciones sustanciales para la identidad tipo de los estados psicológicos. Hemos argumentado que ninguna de las principales teorías actuales parece tener éxito en esta tarea. Podría resultar aconsejable, por lo tanto, reconsiderar toda la empresa.

Supongamos que alguien quisiera saber cuáles son los criterios para identidad tipo de las entidades físicas fundamentales. Quizá la mejor forma de responder a esto es decir que dos entidades son de tipo idéntico si no difieren con respecto a ninguna magnitud física fundamental. Así, hasta donde sabemos, las condiciones para la identificación tipo de las partículas físicas elementales no refieren a su distancia del Polo Norte, aunque sí refieren a su carga. Pero observemos que esto es simplemente una consecuencia del hecho de que no existe ninguna ley física fundamental que opere sobre las entidades en función de su distancia del Polo Norte, y que *sí existen* leyes físicas fundamentales que operan sobre entidades en función de su carga.

Se podría decir que la condición básica para la identidad tipo en la ciencia es que posibilite la articulación del dominio de las leyes. Este principio se sostiene en cada nivel de descripción científica. Así, lo que resulta *pertinente* para la pregunta de si dos entidades de un nivel son de tipo distinto es el carácter de las leyes que operan sobre entidades en ese nivel. Pero si esto es lo que en general ocurre, parecería entonces que las condiciones sustanciales para la identidad tipo de los estados psicológicos se fijarán con referencia a las leyes psicológicas (y quizá neurológicas) que operan sobre esos estados, y no de ninguna otra manera.

A la luz de estas observaciones, podemos obtener una visión más clara de cuál ha sido el error con las distintas clases de teorías filosóficas que hemos venido rechazando. Por ejemplo, cabría pensar que el conductismo entraña un intento por identificar a nivel tipo estados psicológicos sólo por referencia a las leyes, cualesquiera que sean, que determinan sus efectos *conductuales*. Pero, incluso a primera vista, esto parecería un error, ya que debe haber leyes que gobiernen la interacción de los estados psicológicos, y no hay ninguna razón para creer (y muchas razones para no creer) que los estados psicológicos que se comportan uniformemente según las leyes de la primera clase se suelen comportar uniformemente según las leyes de la segunda clase.

De modo análogo, el error del fisicalismo ha sido suponer que los estados psicológicos que son distintos en su conduc-

ta según las leyes neurológicas son *ipso facto* distintos en su conducta según las leyes psicológicas. Sin embargo, todas las probabilidades indican que estados neurológicos distintos pueden ser funcionalmente idénticos. Esto es, satisfacer los criterios para la distinción tipo de estados neurológicos probablemente no garantice la satisfacción de los criterios para la distinción tipo de estados psicológicos, ni viceversa.

En suma, el problema fundamental con el conductismo y el fisicalismo radica en que la identidad tipo se determina en relación con, en el mejor de los casos, un subconjunto de leyes, que se debe suponer, se aplican a los estados psicológicos. La única justificación de esta restricción parece encontrarse en las inclinaciones reduccionistas de estas posturas. Una vez que se ha cuestionado el reduccionismo, podemos ver que las exigencias nomológicas para la identificación tipo de los estados psicológicos son probablemente sumamente complejas y variadas. Incluso lo poco que sabemos sobre las leyes psicológicas hacen parecer poco plausible que éstas reconozcan límites de tipos entre los estados psicológicos en los sitios donde los fisicalistas o los conductistas quieren demarcarlos.

La falla básica de la TIEF es, en ciertos aspectos, análoga a la del conductismo y el fisicalismo. Desde luego, la TIEF no es reduccionista, ni siquiera en espíritu, y en tanto se trata de una especie de funcionalismo, nos invita a identificar los estados psicológicos en tipos haciendo referencia a sus conexiones nomológicas con las entradas sensoriales, las salidas conductuales y las relaciones entre sí. Pero la TIEF pretende fijar una restricción adicional a la identidad tipo: a saber, que se puede establecer una correspondencia (de hecho, una identificación) entre los estados psicológicos de un organismo y los estados de tabla de máquina especificados por la *descripción* óptima del organismo. Hemos argumentado que ésta es, de hecho, una restricción sustancial, y que no puede ser satisfecha.

Lo que aparentemente queda de la TIEF es lo siguiente. Puede ser a la vez cierto e importante que los organismos sean autómatas probabilísticos. Pero incluso si es cierto e impor-

tante, el hecho de que los organismos sean autómatas probabilísticos parece tener poco o nada que ver con las condiciones para la identidad tipo de sus estados psicológicos.

[*Traducción de Ana Isabel Stellino*]

BIBLIOGRAFÍA

Butler, R.J. (comp.), 1965, *Analytical Philosophy. Second Series*, Blackwell, Oxford.

Capitan, W.H. y D.D. Merrill, 1967, *Art, Mind and Religion*, Pittsburgh University Press, Pittsburgh.

Davidson, D., 1970, "Mental Events", en Foster y Swanson 1970, pp. 79–101. [Versión en castellano: "Sucesos mentales" en D. Davidson, *Ensayos sobre acciones y sucesos*, trad. O. Hansberg, J.A. Robles y M.M. Valdés, Crítica/Instituto de Investigaciones Filosóficas–UNAM, Barcelona, 1995, pp. 263–287.]

———, 1963, "Actions, Reasons and Causes", *Journal of Philosophy*, vol. 60, no. 23, pp. 685–699. [Versión en castellano: "Acciones, razones y causas", en D. Davidson, *Ensayos sobre acciones y sucesos*, pp. 17–36.]

Fodor, J.A., 1968, *Psychological Explanation*, Random House, Nueva York. [Versión en castellano: *La explicación psicológica*, trad. J.E. García Albea, Cátedra, Madrid, 1980.]

Foster, L. y J.W. Swanson (comps.), 1970, *Experience and Theory*, University of Massachussetts Press, Amherst.

O'Connor, J. (comp.), 1969, *Modern Materialism: Readings on Mind-Body Identity*, Harcourt, Brace and World, Nueva York.

Putnam, H., 1970, "On Properties", en Rescher 1969, pp. 235–254. [Versión en castellano: *De las propiedades*, trad. M. Gorostiza, Instituto de Investigaciones Filosóficas–UNAM, México, 1983 (Cuadernos de Crítica, 19).]

———, 1967, "Psychological Predicates", en Capitan y Merrill 1967, pp. 37–48.

———, 1966, "The Mental Life of Some Machines", en O'Connor 1969, pp. 263–281. [Versión en castellano: *La vida mental de algunas máquinas*, trad. M. Gorostiza, Instituto de Investigaciones Filosóficas–UNAM, México, 1981 (Cuadernos de Crítica, 17).]

Putnam, H., 1965, "Brains and Behavior", en Butler 1965, pp. 1-19. [Versión en castellano: *Cerebro y conducta*, trad. R. Amieva, Instituto de Investigaciones Filosóficas-UNAM, México, 1983 (Cuadernos de Crítica, 23).]
Rescher, N. (comp.), 1969, *Essays in Honor of C.G. Hempel*, Reidel, Dordrecht.

FUNCIONALISMO Y *QUALIA*

SYDNEY SHOEMAKER

I

En su artículo "Lo que no son los estados psicológicos", N.J. Block y J.A. Fodor plantean una serie de objeciones a la "teoría de identidad de los estados funcionales" (TIEF), la cual postula que "para cualquier organismo que satisfaga predicados psicológicos, existe una *descripción* única óptima, conforme a la cual cada estado psicológico del organismo es idéntico a uno de sus estados de tabla de máquina relativos a esa descripción".[1] La TIEF es una versión del "funcionalismo" que ellos caracterizan como la doctrina más general de que "las condiciones para la identidad tipo de los estados psicológicos refieren sólo a sus relaciones con las entradas, las salidas y a las relaciones entre sí".[2] Block y Fodor consideran que la mayoría de las objeciones que plantean están dirigidas sólo a la TIEF, y no al funcionalismo considerado en términos más amplios. No me ocuparé aquí de estas objeciones. Pero plantean una objeción que, dicen, "podría tomarse [...] como si demostrara que los estados psicológicos no pueden definirse *en lo absoluto* en términos funcionalistas, y que no pueden corresponder a *ninguna* propiedad definible en términos de autómatas abstractos".[3] En pocas pala-

[1] Block y Fodor 1972, p. 165 [p. 161; incluido en esta antología, pp. 155–181; en adelante las páginas que corresponden a esta versión en español del artículo aparecerán entre corchetes].

[2] Block y Fodor 1972, p. 173 [p. 171].

[3] Block y Fodor 1972, pp. 173–174 [p. 171].

bras, la objeción es que la manera de "establecer una identidad tipo" de los estados psicológicos propuesta por la TIEF, y por el funcionalismo en general, "no logra dar cabida a un rasgo de por lo menos algunos de esos estados, el cual es crucial para determinar su tipo: a saber, su carácter 'cualitativo' ".[4]

Block y Fodor sólo dedican un par de páginas a esta objeción, y la plantean en términos bastante tentativos; así que es muy probable que la extensión con que la discuto aquí sea desproporcionada frente a la importancia que ellos le otorgan. Pero han dado una formulación concisa y vívida de una objeción que se siente y se expresa en conversación con mayor frecuencia de lo que aparece impresa, y que me parece plantea algunas cuestiones fundamentales. Otros filósofos han planteado casi la misma objeción al decir que el funcionalismo (o el conductismo, o el materialismo, o las teorías "causales" de la mente –pues la objeción se ha hecho a todos ellos) no puede dar cuenta del componente de "sensación cruda" de los estados mentales, o de su carácter "interno" o "fenomenológico". Mi interés principal aquí no es si esta objeción resulta fatal o no para TIEF; si entiendo esa teoría de manera correcta, queda suficientemente refutada por las otras objeciones que Block y Fodor plantean contra ella. Pero la manera en que ellos caracterizan el funcionalismo "en sentido amplio", es, aunque vaga, una idea que muchos filósofos, incluyéndome, encuentran atractiva; y me parece que vale la pena considerar si puede ser defendida o no de esta objeción.

Seguiré a Block y Fodor al hablar de los estados mentales (o más bien, de algunos estados mentales) como poseedores de "carácter cualitativo" o "contenido cualitativo". Espero que, en la discusión que sigue, se ponga de manifiesto que esto no me compromete con nada que a un lúcido opositor de los "objetos privados" o del "lenguaje privado" le parezca objetable.

II

Block y Fodor desarrollan su objeción en dos etapas. La primera es la que llaman el "argumento de los *qualia* invertidos", y la segunda puede llamarse el "argumento de los *qualia* ausentes".

[4] Block y Fodor 1972, p. 172 [p. 169].

Puesto que no los convencen los conocidos argumentos "verificacionistas" contra la coherencia conceptual de la "hipótesis del espectro invertido", Block y Fodor se inclinan a pensar que los casos de "*qualia* invertidos" pueden ser posibles. Piensan que habría una inversión de los *qualia* (seguramente, un caso extremo de ésta) si fuera verdad que "cada persona tiene, de hecho, *qualia* ligeramente distintos (o, mejor aún, *qualia* extremadamente distintos) cuando está en cualquier estado de tabla de máquina que se supone es idéntico al dolor".[5] La posibilidad de esto resulta incompatible con el funcionalismo bajo el supuesto plausible de que "no habría ningún caso del tipo 'estado de dolor' a menos que se sintiera como dolor [...] aun cuando estuviera relacionado con todos los otros estados psicológicos del organismo de cualesquiera que sean las formas en que los dolores están relacionados".[6]

Block y Fodor no consideran que la posibilidad de la inversión de los *qualia* constituya por sí misma una objeción decisiva al funcionalismo, pues piensan que el funcionalista es libre de negar el supuesto, plausible a primera vista, de que los dolores deben ser cualitativamente similares (y, seguramente, el supuesto relacionado de que todo lo cualitativamente idéntico a un dolor es él mismo un dolor).[7] Si la inversión de los *qualia* de

[5] Block y Fodor 1972, p. 173 [p. 170].

[6] Block y Fodor 1972, p. 172 [p. 169]. Vale la pena observar que este supuesto, o uno muy parecido, desempeña un papel crucial en los recientes argumentos de Saul Kripke contra la teoría de la identidad psicofísica; Kripke la expresa diciendo que el dolor "no es seleccionado por una de sus propiedades accidentales; más bien es seleccionado por la propiedad de ser dolor mismo, por su cualidad fenomenológica inmediata. Así, el dolor [...] no se designa rígidamente por 'dolor', pero la referencia del designador se determina mediante una propiedad esencial del referente". (Kripke 1972, p. 340).

[7] Block y Fodor mencionan otra forma, además de la citada en el texto, en que un funcionalista podría tratar de responder al argumento de los *qualia* invertidos; podría sostener que "aunque los *qualia* invertidos, *si se dieran*, ofrecerían contraejemplos de su teoría, es imposible, nomológicamente, que estados psicológicos funcionalmente idénticos sean cualitativamente distintos" (p. 172) [p. 170]. La idea en este caso debe ser que la simple posibilidad lógica, o conceptual, de la inversión de los *qualia* no es incompatible con el funcionalismo. Parecería, no obstante, que si el que los *qualia* invertidos ocurran realmente ha de proporcionar contraejemplos al funcionalismo (según concede la respuesta imaginada), entonces la mera posibilidad lógica de los *qualia*

hecho se produjera en el caso del dolor (es decir, si un estado funcionalmente idéntico a un dolor difiriese de éste en su carácter cualitativo), entonces, dicen ellos, "podría ser razonable afirmar que el carácter de los *qualia* de un organismo carece de pertinencia con respecto a si tiene un dolor o (en términos equivalentes) que los dolores se sienten de manera muy distinta en organismos diferentes".[8] Tal idea, de hecho, no es insólita. Según Don Locke, "El que una sensación sea una sensación de dolor no es una cuestión sobre cómo se siente, sino una cuestión sobre ser del tipo de cosa que es causada por un daño corporal y conducente a una conducta de dolor."[9] Y Alan Donagan le ha atribuido a Wittgenstein la idea de que

> tú y yo decimos correctamente que tenemos la misma sensación, por ejemplo, dolor de muelas, si ambos padecemos algo horrible que expresaríamos de un modo natural sujetándonos y frotándonos la mandíbula, haciendo ciertos tipos de muecas, y cosas semejantes. El que el carácter interno de lo que se expresa de estas maneras sea o no el mismo para ti que para mí carece de pertinencia para el significado del término "dolor de muelas".[10]

Pero aunque Block y Fodor no descartan esta respuesta al argumento de los *qualia* invertidos como un error obvio, consideran que posiblemente abre la puerta a un argumento mucho más perjudicial para el funcionalismo, a saber, el argumento de los *qualia ausentes*. Su idea quizá es que, una vez admitido que un estado funcional dado puede existir sin tener un "contenido cualitativo" dado, resultará difícil negar la posibilidad de que pueda existir sin tener absolutamente ningún contenido (o carácter) cualitativo. En cualquier caso, continúan diciendo que:

invertidos es incompatible con el funcionalismo; el dolor no puede ser *idéntico* a un estado funcional dado si hay un mundo posible, incluso un mundo lógica pero no nomológicamente posible, en el que el estado funcional existe sin que exista el dolor, o viceversa. (Con respecto a la afirmación general sobre la identidad a la que aquí se está aludiendo, a saber, que si *a* y *b* son idénticos deben ser idénticos en cualquier mundo lógicamente posible en el cual exista uno u otro de ellos, véanse Kripke 1972 y 1971.)

[8] Block y Fodor 1972, p. 173 [p. 170].

[9] Locke 1968, p. 101.

[10] Donagan 1966.

> Hasta donde sabemos ahora, puede ser nomológicamente posible que dos estados psicológicos sean funcionalmente idénticos (es decir, que estén relacionados idénticamente con las entradas, las salidas y los estados sucesores), aun cuando sólo uno de los estados posea un contenido cualitativo. En este caso, la TIEF nos obligaría a decir que un organismo puede tener un dolor aunque no esté sintiendo *absolutamente nada*, y esta consecuencia parece totalmente inaceptable.[11]

Y si son posibles los casos de "*qualia* ausentes", es decir, si un estado puede ser funcionalmente idéntico a un estado que tiene un carácter cualitativo sin tener él mismo un contenido cualitativo, entonces no sólo la TIEF, sino también el funcionalismo en sentido amplio, parecerían ser insostenibles.

III

Si los estados mentales pueden ser iguales o distintos en cuanto a su "carácter cualitativo", deberíamos poder hablar de una clase de estados, llamémoslos "estados cualitativos", cuyas "condiciones de identidad tipo" se pudieran especificar en términos de la noción de similitud cualitativa (o "fenomenológica"). Para cada carácter cualitativo determinado que puede tener un estado, hay (esto es, podemos definir) un estado cualitativo determinado que una persona tiene sólo en caso de que tenga un estado con precisamente ese carácter cualitativo. Por ejemplo, hay un estado cualitativo que alguien tiene sólo en caso de que tenga una sensación que se siente de la misma manera en que se sintió mi dolor de cabeza más reciente. Ahora bien, los estados cualitativos serán ellos mismos estados "mentales" o "psicológicos". Y esto pone en entredicho la sugerencia de Block y Fodor en el sentido de que un funcionalista podía responder al "argumento de los *qualia* invertidos" sosteniendo que "el carácter de los *qualia* de un organismo carece de pertinencia con respecto a si tiene un dolor". Si los estados mentales incluyen a los estados cualitativos, lo que tal funcionalista dice

[11] Block y Fodor 1972, p. 173 [p. 171].

acerca del dolor no podría decirse de los estados mentales en general, ya que sería contradictorio decir que el carácter de los *qualia* de un organismo no es pertinente para la cuestión sobre cuáles estados cualitativos tiene. Y, desde luego, si los estados cualitativos mismos pudieran definirse funcionalmente, entonces la posibilidad de la inversión de los *qualia* no representaría dificultad alguna para el funcionalismo, y el funcionalista no tendría necesidad de negar contraintuitivamente que el carácter de los *qualia* de un organismo es pertinente para si tiene un dolor. Pero si, como parecen suponer Block y Fodor, los estados cualitativos no pueden definirse en términos funcionales, hay entonces una clase de estados mentales, a saber, los estados cualitativos mismos, que no puede definirse funcionalmente.

Esto plantea algunas preguntas a las que volveré en las siguientes secciones, a saber: (a) ¿En qué sentido no son funcionalmente definibles los estados cualitativos (o en qué sentido no pueden definirse funcionalmente si la inversión de los *qualia* es posible)? Y (b) el que sean funcionalmente indefinibles (en cualquier sentido en que lo sean), ¿inflige un grave daño al funcionalismo? Como veremos en el resto de la presente sección, esta pregunta también surge al considerar la presunta posibilidad de los "*qualia* ausentes".

Podemos establecer la imposibilidad de que se den casos de "*qualia* ausentes" si logramos demostrar que, si un estado es funcionalmente idéntico a un estado que posee un contenido cualitativo, entonces él mismo debe poseer un contenido cualitativo. Se podría tratar de hacer esto interpretando la noción de identidad funcional de manera tal que los estados cualitativos queden incluidos entre los "otros estados psicológicos" en relación con los cuales, junto con las entradas y las salidas, la "identidad tipo" de un estado psicológico dado ha de definirse. Así, uno podría argumentar que si un estado psicológico dado posee cierto carácter cualitativo, esto implica que se encuentra en alguna relación determinada con algún estado cualitativo particular (a saber, el estado cualitativo en el que se halla una persona sólo en caso de que esté en un estado que tenga ese carácter cualitativo), y que cualquier estado funcionalmente

idéntico a él debe encontrarse en la misma relación con ese estado cualitativo, y por lo tanto debe tener el mismo carácter cualitativo.[12] Pero este argumento no es muy convincente. Una objeción a éste es que, como los estados cualitativos mismos no pueden definirse en términos funcionales (suponiendo la posibilidad de los *qualia invertidos*), no es legítimo incluirlos entre los estados psicológicos con referencia a los cuales se definen funcionalmente otros estados psicológicos, o en términos de los cuales se define la "identidad funcional". Volveré a esta objeción más adelante, ya que constituye también una objeción *prima facie* al argumento más plausible que presentaré a continuación. Otra objeción consiste en que la relación que un estado guarda con un estado cualitativo, al tener el "carácter cualitativo" correspondiente a ese estado cualitativo, no se parece en lo absoluto a una relación causal y, por lo tanto, no es la clase de relación en términos de la cual pueda definirse funcionalmente un estado psicológico. Pero el argumento que presentaré enseguida no está expuesto a esta objeción, y me parece que sí demuestra que, conforme a cualquier interpretación plausible de la noción de identidad funcional, un estado no puede ser funcionalmente idéntico a un estado que posee un carácter cualitativo sin tener él mismo un carácter cualitativo.

Una forma importante en la cual los dolores se relacionan con otros estados psicológicos consiste en que dan origen, en las circunstancias apropiadas, a una conciencia introspectiva de sí mismos como poseedores de ciertos caracteres cualitativos, esto es, como si se sientieran de ciertas maneras. Supondré que el

[12] ¿Cuál es precisamente la relación que un estado debe guardar con un estado cualitativo a fin de tener el carácter cualitativo correspondiente a ese estado? No puede ser, en los casos que nos ocupan, la relación de identidad (esto sólo permitiría a los estados cualitativos tener carácter cualitativo, y no nos permitiría hablar del carácter cualitativo de los estados cuyas condiciones de "identidad tipo" se dan en términos funcionales). Y es de suponerse que debe ser algo más fuerte que la relación "está acompañada por" o "está coinstanciada con". Lo mejor que puedo hacer es decir que un caso particular de un estado S tiene el carácter cualitativo correspondiente al estado cualitativo Q si, en la ocasión en cuestión, el que se dé un caso (la instanciación) de S esencialmente involucra el que se dé un caso (la instanciación) de Q. Posiblemente, pero no estoy seguro de esto, podríamos fortalecer esto, y hacerlo menos vago, diciendo que en tales ocasiones el caso de S es un caso de Q.

significado de esto puede desarrollarse parcialmente diciendo que tener dolor típicamente da origen, dadas las circunstancias apropiadas, a lo que llamaré una "creencia cualitativa", esto es, una creencia de que uno se siente de cierta manera (o en términos más abstractos, que uno se encuentra en un estado que tiene cierto carácter cualitativo, o, dicho también de otro modo, que uno está en cierto estado cualitativo). Cualquier estado funcionalmente idéntico a un estado de dolor compartirá con el estado de dolor no sólo (1) su tendencia a influir de ciertas maneras en la conducta manifiesta, y (2) su tendencia a producir en la persona la creencia de que hay algo orgánicamente mal en ella (por ejemplo, que se ha cortado o quemado), sino también (3) su tendencia a producir creencias cualitativas en la persona, esto es, a hacerle pensar que tiene un dolor con cierto carácter cualitativo (el cual le desagrada). De acuerdo con el "argumento de los *qualia* ausentes", tal estado puede carecer, no obstante, de carácter cualitativo, y por lo tanto, no ser dolor. Consideremos si esto es plausible.

Suponiendo que tales casos de "*qualia* ausentes" son posibles, ¿de qué manera podríamos detectar un caso semejante si llegara a producirse? Y, ¿con qué derecho cada uno de nosotros rechaza la sugerencia de que quizá su propio caso es uno de esos casos y que uno mismo carece de estados con carácter cualitativo? En efecto, ¿con qué derecho rechazamos la sugerencia de que quizá nadie experimenta jamás sensaciones (u otros estados que poseen carácter cualitativo)? Es, desde luego, una idea conocida que la conducta ofrece evidencia no concluyente acerca de qué carácter cualitativo tienen, si acaso lo tienen, los estados mentales de un hombre. Pero lo que normalmente subyace a esto es la idea de que el hombre mismo tiene un acceso más "directo" a este carácter cualitativo del que jamás podría ofrecer la conducta, a saber, la introspección. Y la introspección, independientemente de qué más pueda ser, es el vínculo entre los estados mentales de un hombre y sus creencias acerca de esos estados (o su conocimiento o conciencia de ellos). Así, una forma de expresar nuestra interrogante es preguntar si hay algo que pueda servir como evidencia (para cualquier persona) de que alguien no sufre un dolor, dado que de los estados en que se encuentra, más las leyes psicológicas que son verdade-

ras de él (las leyes que describen las relaciones de sus estados entre sí y con entradas y salidas), se sigue que la totalidad de la evidencia conductual posible *más* la totalidad de la evidencia introspectiva posible apunta claramente a la conclusión de que tiene un dolor. No veo cómo algo pueda servir como tal. Sin duda, podemos imaginar (quizá) que unos "cerebroscopios" revelan que la persona no está en el estado neurofisiológico en el que nosotros nos encontramos siempre que tenemos (según pensamos) un dolor. Pero esto no hace más que suscitar la pregunta de sobre qué bases podemos decir que *nosotros* tenemos un dolor genuino (esto es, un estado con un carácter cualitativo y que además desempeña el papel funcional apropiado en sus relaciones con entradas, salidas y otros estados psicológicos). Aquí parece que si la evidencia conductual e introspectiva no es suficiente, nada podría serlo. Pero si es suficiente en el caso de nosotros, lo es en el caso de nuestro hombre hipotético. En cualquier caso, si se da por sentado que el estado de un hombre es funcionalmente idéntico a un estado que en nosotros es dolor, es difícil ver cómo una diferencia fisiológica entre él y nosotros podría servir de modo alguno como evidencia de que sus estados carecen de carácter cualitativo; pues si algo puede servir como evidencia para nosotros acerca de su estado psicológico, la evidencia de que su estado es funcionalmente equivalente al nuestro es *ipso facto* evidencia de que cualquier diferencia fisiológica entre nosotros y él carece de pertinencia para la cuestión de si, aunque no de cómo, el estado de dolor se realiza en él.

Sostener que es lógicamente posible (o, peor, nomológicamente posible) que un estado carente de carácter cualitativo deba ser funcionalmente idéntico a un estado con carácter cualitativo es hacer que el carácter cualitativo no sea pertinente tanto para lo que podemos considerar que sabemos al saber de los estados mentales de otros, como también para lo que podemos considerar que sabemos al saber de nuestros propios estados mentales. No podría haber (desde este punto de vista) ningún efecto físico posible de un estado a partir del cual pudiéramos argumentar, por medio de una "inferencia a la mejor explicación", que posee un carácter cualitativo; pues si lo hubiera, podríamos ofrecer por lo menos una caracterización

funcional parcial de la posesión de carácter cualitativo diciendo que tiende a dar origen, en tales y cuales circunstancias, a esos efectos físicos, y no podríamos admitir que un estado que carece de carácter cualitativo sea funcionalmente idéntico a otro que sí lo posee. Y por las razones que ya hemos dado, si fueran posibles casos de "*qualia* ausentes", el carácter cualitativo sería necesariamente inaccesible a la introspección. Si el carácter cualitativo careciera de pertinencia en este sentido para todo conocimiento de las mentes, para el autoconocimiento así como para el conocimiento de otros, no sería en modo alguno "inaceptable", sino más bien sería sólo una cuestión de sentido común negar que los dolores deben tener carácter cualitativo. Pero, desde luego, es absurdo suponer que las personas ordinarias están hablando de algo que, en principio, es incognoscible por todos cuando hablan acerca de cómo se sienten, o de cómo se ven, se oyen, huelen, etc., las cosas para ellos. (En efecto, de la misma forma en que una teoría causal del conocimiento implicaría que los estados o los rasgos que son independientes de los poderes causales de las cosas que ellos caracterizan serían en principio incognoscibles, una teoría causal de la referencia implicaría que tales estados y rasgos son, en principio, innombrables e inaccesibles a la referencia.) Y si, para volver a la sensatez, consideramos que el carácter cualitativo es algo que puede conocerse de las mismas maneras en que consideramos que son cognoscibles las sensaciones humanas (como mínimo, si pueden conocerse introspectivamente), entonces no es posible, ni siquiera lógicamente posible, que un estado que carece de carácter cualitativo sea funcionalmente idéntico a un estado que sí lo posee.

Éste no es un argumento "verificacionista". No supone ninguna conexión general entre significatividad y verificabilidad (o cognoscibilidad). Lo que sí asume es que, si ha de haber una razón para suponer (como lo hace el "argumento de los *qualia* ausentes") que es esencial para el dolor y otros estados mentales poseer "carácter cualitativo", entonces debemos tomar la expresión "carácter cualitativo" como refiriendo a algo que es cognoscible por lo menos en algunas de las formas en las que consideramos cognoscibles para los dolores (los nuestros y los de otros). También asume que si pudiera haber un rasgo de

algún estado mental que fuera enteramente independiente de los poderes causales de ese estado (es decir, que fuera tal que su presencia o su ausencia no afectara las tendencias del estado a producir otros estados, y demás), y así que no fuera pertinente para su "identidad funcional", entonces ese rasgo sería totalmente incognoscible (si se quiere, esto supone una teoría causal del conocimiento).

Contra este argumento, al igual que contra uno anterior, se puede objetar que los otros estados psicológicos, en relación con los cuales (junto con las entradas y salidas) se define funcionalmente un estado psicológico dado, no deben incluir ningún estado que no pueda ser en sí mismo definido funcionalmente. Pues podría decirse que los estados que he llamado "creencias cualitativas" no pueden definirse funcionalmente como tampoco es posible hacerlo con los estados cualitativos mismos. La relación más importante de estos estados con otros estados parecería ser su relación con los estados cualitativos que, típicamente, les dan origen, pero (según sigue el argumento) estos últimos no pueden ser definidos funcionalmente y, por ende, no es legítimo hacer referencia a ellos en las definiciones funcionales de los primeros. Además (recordando que aquí no se está cuestionando la posibilidad de casos de *qualia invertidos*), parece plausible suponer que si dos personas difiriesen con respecto al carácter cualitativo de sus dolores, pero en forma tal que la diferencia no se manifestara en ninguna conducta posible, también diferirían en sus creencias cualitativas, y esta última diferencia sería tal que su existencia tampoco podría manifestarse en ninguna conducta posible. Y si esto es posible, parece que hay tantas razones para negar que las creencias cualitativas puedan definirse funcionalmente como para negar que los estados cualitativos puedan definirse funcionalmente.

Esta objeción no menciona una cuestión importante que está implícita en mi argumento, a saber, que no podemos negar, sin quedar comprometidos con un intolerable escepticismo acerca de los dolores de los demás, que el que una persona diga que siente un dolor agudo es una evidencia sólida de que está en tal o cual estado cualitativo, y lo es porque el que una persona diga esto es, normalmente, un *efecto* de estar en un estado que posee carácter cualitativo —y esto por sí mismo sugiere fuerte-

mente que, si un estado mental de una persona posee carácter cualitativo, y si cierto estado de otra persona, que en lo demás es similar al anterior, carece de carácter cualitativo, entonces los estados difieren en las maneras en que tienden a influir en la conducta ("la salida") y, por lo tanto, difieren funcionalmente. De cualquier forma, la posibilidad de los "*qualia* invertidos" sí parece implicar que los estados cualitativos, y por ende las creencias cualitativas, no pueden ser definidos funcionalmente. Para determinar si esto es compatible con el funcionalismo, y si socava el argumento expuesto antes, necesitamos considerar en qué sentido es verdad que los estados cualitativos (y las creencias cualitativas) no son funcionalmente definibles, y qué límites hay sobre la manera en que la referencia a estados mentales que no son funcionalmente definibles puede introducirse en las definiciones funcionales de otros estados mentales.

Para considerar estas cuestiones quisiera cambiar de ejemplo, y pasar de nuestro estudio del caso del dolor al de la experiencia visual. Hay dos razones por las que este cambio resulta deseable. En primer lugar, la posibilidad de la "inversión del espectro" (cuando la experiencia que tiene una persona de los colores difiere sistemáticamente, en su carácter cualitativo o fenomenológico, de la experiencia que otra persona tiene de los mismos colores) me parece mucho menos problemática que la posibilidad de la "inversión de los *qualia*" en el caso del dolor (cuando diferentes personas sienten un dolor en forma radicalmente distinta). En segundo lugar, y relacionado con esto, es mucho más fácil distinguir ver azul (por ejemplo) de su carácter cualitativo que distinguir el dolor de su carácter cualitativo, y en consecuencia es mucho más fácil considerar de qué manera la referencia a los estados cualitativos podría introducirse en una explicación funcional de ver los colores que considerar de qué manera la referencia a esos estados podría introducirse en una explicación funcional del dolor.

IV

Si veo algo, se me aparece de algún modo, y la manera en que se ve se asemeja y difiere, en diversos grados y varios aspectos, de las maneras en que otras cosas se me aparecen o se me

han aparecido en otras ocasiones. Es debido a que las similitudes y las diferencias entre estas "maneras de aparecerse a" se correlacionan de modos sistemáticos con las similitudes y diferencias entre los objetos ya vistos que somos capaces de ver estos objetos y sus propiedades en virtud de las cuales se dan las similitudes y las diferencias.[13] Considero que "aparecerse a" de cierta manera, por ejemplo, que las cosas se le aparezcan a alguien del mismo modo en que se me aparecen ahora al mirar por la ventana, es un estado cualitativo. Así, ver esencialmente conlleva el que ocurran estados cualitativos. Además, la referencia a estos estados cualitativos entra en lo que se asemeja mucho a una explicación funcional del ver. Pues parecería que lo que significa decir que alguien ve algo como siendo azul es más o menos lo siguiente:

> *S* ve algo como siendo azul si y sólo si (1) *S* posee un repertorio de estados cualitativos que incluye un conjunto de estados *K* que están asociados con los colores de los objetos en forma tal que (a) la estimulación visual causada por un objeto de cierto color en "condiciones estándar" produce en la persona el estado cualitativo asociado, y (b) los grados de similitud "cualitativa" o "fenomenológica" entre los estados de *K* corresponden a los grados de similitud entre los colores asociados; y (2) la persona *S* (a) está actualmente en el estado cualitativo asociado con el color azul, (b) lo está como resultado de la estimulación visual causada por algo azul, y (c) cree, debido a (a) y (b), que hay algo azul ante ella.[14]

Debo ahora matizar la afirmación de que "aparecerse a" de cierta manera es un estado cualitativo. Si se le pide a un hombre que describa la manera en que se le aparecen a él o, en términos

[13] Tomo la terminología "aparecerse a" de Roderick Chisholm; véase su 1965, esp. la p. 480, n. 6. "Se aparece a" uno tanto en casos de percepción verídica como en casos de ilusión y alucinación, y se le puede "aparecer a" de la misma manera en todas estas clases de casos. La locución técnica aparecerse-azul-a se utiliza en el texto como abreviatura de la locución "ve o le parece ver algo azul" (en una acepción "no epistémica" de esa locución).

[14] Como análisis, esto no servirá muy bien. Puedo ver una cosa como siendo azul aunque se vea verde (es decir, incluso si mi estado cualitativo visual es el asociado con el verde), si me han "informado" que en estas circunstancias las cosas azules se ven verdes.

más naturales, cómo se ven las cosas para él, podría decir, entre otras cosas, que cierto objeto le parece azul, o que a él le parece como que está viendo algo azul, o (si se trata de un filósofo que utiliza el "lenguaje de aparecerse") que "se le aparece azul a" él. Y es natural convertir en una condición de que a alguien se le aparezca azul que esa persona se encuentre en el estado cualitativo que, en ese momento y en ella, está asociado con la estimulación visual causada por las cosas azules; es decir, es natural ofrecer un análisis de "a *S* se le aparece azul" que sea igual al análisis presentado anteriormente de "*S* ve algo como siendo azul", excepto que se suprimen las cláusulas (b) y (c) de la condición (2). Pero si hacemos esto, entonces aparecerse-azul-a no será en sí mismo un estado cualitativo. O en todo caso, esto será así si la inversión del espectro es posible. Podríamos resumir la situación diciendo que aparecerse-azul-a es, en el análisis propuesto, un estado funcional cuya caracterización funcional le exige tener siempre algún carácter cualitativo (u otro), pero sin exigirle que tenga el mismo carácter cualitativo en personas distintas (suponiendo la posibilidad de una inversión intersubjetiva del espectro) o en la misma persona en momentos diferentes (suponiendo la posibilidad de una inversión intrasubjetiva del espectro). Pero esto vuelve a plantear la cuestión de si los estados cualitativos son ellos mismos funcionalmente definibles y, en caso de no serlo, si es legítimo hacer referencia a ellos en las caracterizaciones funcionales de otros estados mentales.

La expresión "aparecerse-azul-a" creo que podría tener un uso según el cual significara un estado cualitativo. Yo podría "fijar la referencia" de esta expresión estipulando que refiere a (o, puesto que se trata de un predicado y no de un término singular, que predica o adscribe) ese estado cualitativo que en el tiempo presente (abril de 1974) está asociado en mí con el ver de cosas azules.[15] Entendiendo la expresión de este modo,

[15] Estoy tomando de Saul Kripke la noción de "fijar la referencia" y la noción de un "designador rígido" que empleo más adelante; véase Kripke 1972, pp. 269–275 y *passim*. El uso de una descripción definida "el *x* tal que *Fx*" para "fijar la referencia" de un término *T* contrasta con la definición de *T* como equivalente en significado a la descripción definida, es decir, como una abreviatura de ésta; en el primer caso, pero no en el segundo, el enunciado "si *T*

si mañana sufriera una inversión del espectro, dejaría de ser el caso de que normalmente se me aparezca azul a mí cuando veo cosas azules, y podría volverse el caso de que normalmente se me aparezca amarillo en tales ocasiones.[16] (Por contraste, en el sentido "funcional" de "aparecerse-azul-a" esbozado anteriormente, podría ser verdad antes y después de la inversión intrasubjetiva del espectro que normalmente se me aparece azul cuando veo cosas azules, aunque desde luego aparecerse-azul-a tendría, en el momento posterior, el carácter cualitativo que otro estado visual, digamos, aparecerse-amarillo-a, tenía en el momento previo.) No creo que fuera de mucha utilidad tener expresiones que fueran, en este sentido, "designadores rígidos" (o "predicados rígidos") de los *qualia* visuales. Por otro lado, no encuentro razón alguna en principio de por qué no pudiéramos tenerlos. Pero si los tuviéramos, no podrían ser definidos funcionalmente. Tales términos tendrían que ser introducidos por una "fijación de la referencia" kripkeana o (lo que constituye un caso especial de esto) por una definición ostensiva. Sin duda, está la posibilidad teórica de dar una definición verbal de una de estas expresiones utilizando otras expresiones del mismo tipo; de igual forma en que yo podría definir "azul" por medio de una descripción de la forma "el color que no es amarillo, ni rojo, ni verde... etc.", de suerte que podría definir "aparecerse-azul-a" como equivalente a una descripción de la forma "el *quale* de color que no es ni aparecerse amarillo a, ni aparecerse rojo a, ni aparecerse verde a... etc." Pero esto es de poco interés, ya que es obviamente imposible que los nombres (o los predicados)

existe, entonces *T* es el *x* tal que *Fx*" será contingentemente verdadero más que necesariamente verdadero. Una expresión es un designador rígido si designa el mismo objeto en todos los mundos posibles (o en todos los mundos posibles en los que designa algo). De acuerdo con Kripke, los nombres comunes son designadores rígidos, mientras que muchas descripciones definidas no lo son. Cuando una descripción definida se utiliza para introducir un nombre (y por lo tanto un designador rígido), se la emplea para "fijar su referencia" más que para "definirlo" o para darle su "significado".

[16] Mi distinción entre el sentido "funcional" de "aparecerse-azul-a" y un sentido (posible) en el cual designa rígidamente a (o mejor, predica rígidamente) un estado cualitativo es parecida a la que establece Chisholm entre los sentidos "comparativos" y "no comparativos" de expresiones como "se ve azul". Véase Chisholm 1957, cap. 4.

de todos los *qualia* visuales deban definirse de esta manera sin circularidad. Así, suponiendo que el discurso acerca de definir estados funcionales es equivalente al discurso acerca de definir nombres o "designadores rígidos" para los estados cualitativos, parece haber un sentido apropiado en el cual los estados cualitativos no pueden definirse funcionalmente.

Pero lo que parece obligarnos a esta conclusión es la aparente posibilidad de la inversión del espectro. Creo que lo que nos obliga a admitir (si acaso hay algo) la posibilidad de la inversión del espectro es la aparente concebibilidad y detectabilidad de la inversión del espectro *intra*subjetiva. Y si reflexionamos en esto último, veremos, creo, que mientras que no podemos definir estados cualitativos particulares de manera funcional, hay un sentido en el cual podemos definir funcionalmente la *clase* de estados cualitativos –podemos definir funcionalmente las condiciones de identidad de los miembros de esta clase, pues podemos definir funcionalmente las relaciones de similitud y diferencia cualitativa (fenomenológica). Esto es lo que argumentaré en la siguiente sección.

V

Interpretada de cierta manera, la afirmación de que la inversión del espectro es posible implica una afirmación que, hasta donde sé, podría ser empíricamente falsa, a saber, que existe un modo de mapear determinados tonos de color con otros determinados tonos de color, de suerte que (1) todo tono determinado (incluyendo "lodoso" y colores no saturados, además de los colores espectrales puros) se mapee con algún tono determinado, (2) por lo menos algunos de los tonos se mapeen con tonos que son distintos de ellos mismos, (3) el mapeo preserve, para cualquier persona de visión normal, todas las relaciones de "distancia" y "comparación" entre los colores (de modo que si los tonos *a*, *b* y *c* se mapean con los tonos *d*, *e* y *f*, respectivamente, una persona de visión normal haga los mismos juicios de similitud comparativa acerca de *a* en relación con *b* y *c* que acerca de *d* en relación con *e* y *f*), y (4) el mapeo preserve todas nuestras intuiciones, salvo aquellas que están condicionadas empíricamente por el conocimiento de las propiedades de mez-

clado de pigmentos y demás, acerca de cuáles tonos son colores "puros" y cuáles tienen otros colores "en" ellos (de manera que, por ejemplo, si los tonos *a* y *b* se mapean con los tonos de anaranjado y rojo, respectivamente, nos inclinaremos a decir que *a* es menos puro que *b* y quizá que tiene a *b* en él). Pero aun si nuestra experiencia de colores de hecho no es tal que sea posible un mapeo de este tipo, me parece concebible que podría haberlo sido –y esto es lo que importa para los propósitos filosóficos que nos ocupan.[17] Por ejemplo, pienso que sabemos bastante bien cómo sería ver el mundo acromáticamente, esto es, en negro, blanco y los diversos tonos de gris, ya que con frecuencia lo vemos así en fotografías, en el cine y la televisión. Y existe un mapeo obvio de los tonos acromáticos entre sí que satisface las condiciones de la inversión. En la discusión que sigue asumiré, por conveniencia, que tal mapeo es posible para toda la gama de colores, pero no creo que se derive nada esencial de si este supuesto es correcto o no.

Suponiendo que haya ese mapeo (y, un supuesto adicional por conveniencia, que sólo haya uno), llamemos al tono con el que se mapea cada tono el "inverso" de ese tono. Tendremos una inversión *inter*subjetiva del espectro si la manera en que cada tono de color le parece a una persona es la manera en que su inverso le parece a otra persona, o en otras palabras, si para cada tono de color el estado cualitativo asociado en una persona con el ver de ese tono está asociado en otra persona con el ver del inverso de ese tono. Y tendremos una inversión *intra*subjetiva del espectro si se produce un cambio en la manera en que los diversos tonos de color le parecen a una persona, de suerte que cada uno acaba por verse de la misma manera en que antes se veía su inverso.

[17] En un libro (Harrison 1973) de cuya existencia me enteré después de haber escrito este ensayo, Bernard Harrison presenta evidencia empírica contra la posibilidad de lo que yo llamo inversión del espectro. También intenta mostrar, sobre una base *a priori* –y en este punto lo encuentro mucho menos convincente–, que "la maquinaria lingüística y conceptual que rige el nombrar los colores funciona de manera tal que cualquier diferencia en el contenido percibido de las presentaciones de los colores, vistos por hablantes distintos, debe revelarse en diferencias en la forma en que éstos aplican los nombres de los colores, o en los privilegios de su presencia en contextos oracionales que exhiben los nombres de los colores en su discurso" (p. 133).

Lo que más nos llama la atención sobre la inversión del espectro es que, si puede ocurrir *inter*subjetivamente, parecería que no hay modo de determinar si la experiencia de color que tienen dos personas es la misma o si sus espectros de color están invertidos uno respecto del otro. La diferencia sistemática entre experiencias en la que consistiría una inversión intersubjetiva del espectro no estaría abierta, desde luego, a la introspección de nadie. Y no parecería haber modo alguno en el cual estas diferencias pudieran manifestarse en la conducta; la hipótesis de que tu espectro está invertido en relación con el mío y la hipótesis de que nuestra experiencia de color es la misma parecen dar origen a las mismas predicciones acerca de nuestra conducta. Aquí, por supuesto, tenemos en mente el caso hipotético en el que los diversos colores siempre se han visto de una manera para una persona y de manera distinta para otra persona. Y la situación parece muy distinta cuando consideramos el caso de la inversión *intra*subjetiva del espectro. En primer lugar, parece que tal cambio se haría manifiesto a la introspección, o a la introspección más la memoria, de la persona en quien se ha producido. Pero si esto es así, otras personas podrían saber de ello a través de los reportes de esa persona. Además, y esto suele observarse con menos frecuencia, hay conductas no verbales, además de las verbales, que podrían indicar que hubo tal cambio. Si un animal ha sido entrenado para responder de modos específicos a objetos de ciertos colores, y luego empieza espontáneamente a responder de esas maneras ante las cosas que poseen los colores inversos, y si muestra sorpresa cuando sus respuestas ya no reciben las recompensas acostumbradas, esto será sin duda evidencia de que ha sufrido una inversión del espectro. En el caso de una persona, podríamos tener una combinación de esta clase de evidencia y de la evidencia de su propio testimonio.[18]

Si no creyéramos que podemos tener estos tipos de evidencia de una inversión intrasubjetiva del espectro, pienso que no tendríamos ninguna razón en lo absoluto para pensar que cual-

[18] Véase la sección II [de Shoemaker 1982 o del ensayo 15 de Shoemaker 1984]; el ejemplo utilizado aquí se presentó originalmente en una nota al pie en ese ensayo.

quier clase de inversión del espectro, intrasubjetiva o intersubjetiva, es siquiera lógicamente posible. Afirmar que la inversión del espectro es posible pero que es indetectable incluso en el caso intrasubjetivo sería cortar la conexión que, suponemos, existe entre los estados cualitativos y la conciencia introspectiva de ellos (entre ellos y las creencias cualitativas a las que dan origen), y también sus conexiones con las creencias perceptuales acerca del mundo y, a través de estas creencias, con la conducta. Sin duda, se podría definir el término "estado cualitativo" de modo tal que no fuera esencial a los estados cualitativos tener estos tipos de conexiones. Pero entonces no sería en virtud de las similitudes y de las diferencias entre los "estados cualitativos" (en el sentido así definido) que las cosas se ven similares y diferentes para la gente, y la hipótesis de que las personas difieren radicalmente con respecto a los "estados cualitativos" que tienen cuando ven cosas de colores diversos no sería de interés filosófico, y no sería la "hipótesis del espectro invertido" como se entiende habitualmente. De hecho, el supuesto de que podría ocurrir una inversión intrasubjetiva del espectro, pero que no sería posible detectarla, es incoherente en un sentido muy semejante al de la "hipótesis de los *qualia* ausentes", es decir, el supuesto de que estados "funcionalmente idénticos" a estados con contenido cualitativo pudieran ellos mismos carecer de contenido cualitativo. Ninguno de los dos supuestos tiene sentido a menos que las nociones cruciales en ellos estén implícitamente definidas, o redefinidas, de manera tal que el supuesto se vuelva vacuo o carente de interés.

Pero, entonces, ¿qué estamos suponiendo sobre los estados cualitativos, y sobre las relaciones de similitud y diferencia cualitativas o fenomenológicas entre estos estados, cuando suponemos que una inversión del espectro intrapersonal *sería* detectable? En lo que sigue, hablaré de los estados cualitativos caso como "experiencias", y diré que las experiencias son "coconscientes" si son conscientes para una persona al mismo tiempo, donde una experiencia cuenta como consciente para una persona cuando la recuerda de manera correcta, así como cuando está de hecho teniéndola. Una de las cosas que estamos suponiendo, si consideramos la inversión intrasubjetiva del espectro como detectable en las formas que he indicado, es que

cuando las experiencias son co-conscientes las similitudes entre ellas tienden a dar origen a la creencia en la existencia de similitudes objetivas en el mundo físico, a saber, similitudes entre objetos durante cuya percepción se produjeron las experiencias, y las diferencias entre ellas tienden a dar origen a la creencia en la existencia de diferencias objetivas en el mundo. Y estas creencias, a su vez, dan origen (junto con los deseos y otros estados mentales de la persona) a una conducta manifiesta que es apropiada para ellas. Esto explica cómo puede haber una conducta no verbal que sea evidencia de la inversión del espectro; la conducta será la manifestación de creencias erróneas sobre las cosas, las cuales se dan por el hecho de que, en caso de una inversión intrasubjetiva del espectro, las cosas del mismo color producirán, después de la inversión, experiencias cualitativamente distintas a las que producían antes, mientras que las cosas de cada color producirán, después de la inversión, experiencias cualitativamente semejantes a las que producían antes de la inversión las cosas de color distinto.

Pero incluso si, por alguna razón, una víctima de la inversión del espectro no acabara teniendo creencias erróneas acerca de similitudes y diferencias objetivas de esta manera, y actuando con base en ellas, todavía podríamos tener pruebas de que su espectro se había invertido, ya que nos podría decir que ha sucedido. Y al suponer que *él* puede saber de la inversión del espectro en tal caso y así encontrarse en posición de informarnos de ello, estamos suponiendo algo más sobre las relaciones de similitud y diferencia cualitativas, a saber, que cuando se dan entre experiencias co-conscientes, esto tiende a dar origen a una conciencia introspectiva de la existencia de estas relaciones mismas, es decir, tiende a dar origen a "creencias cualitativas" correctas en el sentido de que estas relaciones existen.

Los filósofos que hablan de los estados mentales como poseedores de "criterios" conductuales han dicho en ocasiones que el criterio de la similitud de experiencias es que el sujeto reporte o esté dispuesto a reportar sinceramente que son similares.[19] Si reformulamos esta idea en términos funcionalistas, aparecerá

[19] Véase, por ejemplo, Ginet 1968, p. 9.

como la idea de que aquello que constituye la similitud cualitativa entre experiencias es, al menos en parte, que dan origen, o tienden a dar origen, a que su sujeto tenga una creencia cualitativa de que tal similitud existe y, en virtud de esta creencia, a una disposición a ofrecer reportes verbales al respecto. Pero, como *definición* funcional de la similitud cualitativa, esto sería, desde luego, circular. Si estamos tratando de explicar lo que significa que las experiencias sean similares, no podemos dar ya por entendida, y como lista para usarse en nuestra explicación, la noción de creer que las experiencias son similares.

Pero no habría ninguna circularidad semejante al definir funcionalmente las nociones de similitud y diferencia cualitativa en términos de la primera clase de relación que mencioné, a saber, entre, por un lado, el que las experiencias de una persona sean cualitativamente similares o diferentes de cierta manera y, por el otro, su creencia en la existencia de ciertas clases de similitudes y diferencias objetivas en el mundo, y, finalmente, el que se comporte de determinada manera. Creo que se puede presentar un caso, aunque no intentaré hacerlo aquí, a favor de decir que la tendencia de las experiencias sensoriales a dar origen a una conciencia introspectiva de sí mismas y de sus similitudes y diferencias es, para las criaturas que poseen las capacidades conceptuales de los seres humanos, una consecuencia inevitable de su tendencia a dar origen a una conciencia perceptual de los objetos en el mundo, y de las similitudes y diferencias entre estos objetos. Y mi sugerencia es que lo que hace que una relación entre experiencias sea la relación de similitud cualitativa (fenomenológica) es precisamente que desempeña cierto papel "funcional" en la conciencia perceptual de las similitudes objetivas, a saber, su tendencia a producir creencias perceptuales de que tales similitudes existen. Asimismo, lo que hace que una relación entre experiencias sea la relación de la diferencia cualitativa es que desempeña el papel correspondiente en la conciencia perceptual de las diferencias objetivas.[20]

Esta sugerencia es, desde luego, vaga y esquemática. Pero todo lo que aquí debo sostener es que la afirmación de que po-

[20] Se presentan argumentos adicionales en favor de esta concepción en Shoemaker 1975 o en el ensayo 8 de Shoemaker 1984.

demos dar cuenta funcionalmente de la similitud y la diferencia cualitativas según estos lineamientos no resulta menos plausible que la afirmación de que estados mentales como la creencia y el deseo pueden definirse funcionalmente. Pues lo que busco no es el objetivo ambicioso de demostrar que el funcionalismo ofrece una filosofía de la mente plenamente satisfactoria, sino el propósito mucho más modesto de mostrar que el hecho de que algunos estados mentales posean "carácter cualitativo" no presentan necesariamente ninguna dificultad especial al funcionalista. Y un paso importante para mostrar esto último consiste en mostrar que las nociones de similitud y diferencia cualitativas son tan plausibles como objetos de una definición funcional como otras nociones mentales. Concedí anteriormente que hay un sentido en el cual los estados cualitativos particulares no pueden definirse funcionalmente. Pero debe recordarse que lo que distingue a los estados cualitativos de otras clases de estados mentales es que sus "condiciones de identidad tipo" han de darse en términos de la noción de similitud cualitativa. Al principio de nuestra discusión, especificar las condiciones de identidad en esos términos parecía contrastar marcadamente con su especificación en términos funcionales. Pero este contraste se difumina si, como he sugerido, la propia noción de similitud cualitativa puede definirse en términos funcionales. Y si esto último es así, y por ende las condiciones de identidad de los estados cualitativos pueden especificarse en términos funcionales, no parece inapropiado decir, como lo hice antes, que mientras que los estados cualitativos particulares no pueden definirse funcionalmente, sí puede definirse funcionalmente la *clase* de estados cualitativos.

VI

Retomemos ahora la pregunta de si es legítimo hacer referencia a los estados cualitativos al dar definiciones funcionales de otros tipos de estados mentales.

De acuerdo con una de sus interpretaciones, el funcionalismo en la filosofía de la mente es la doctrina de que los términos mentales o psicológicos son, en principio, eliminables de cierta manera. Si, para simplificar el asunto, consideramos que

nuestro vocabulario mental se compone de nombres de estados y relaciones mentales (más que de predicados que adscriben tales estados y relaciones), la afirmación consistirá en que estos nombres pueden tratarse como sinónimos de descripciones definidas, cada una de las cuales será formulable, en principio, sin utilizar ningún término del vocabulario mental. En efecto, se cuantificará sobre los estados mentales, y en algunos casos serán identificablemente referidos en estas descripciones definidas; pero cuando lo estén, se los caracterizará e identificará, no en términos explícitamente mentalistas, sino en términos de sus relaciones causales y otras relaciones "tópico neutrales" que guardan entre sí y con las entradas y salidas físicas.[21]

Ahora bien, lo que ya he dicho implica que los nombres de los estados cualitativos (si los tuviéramos) no podrían ser definidos como equivalentes a esas descripciones definidas, bajo el supuesto, desde luego, de que es posible la "inversión de los *qualia*". Si el papel causal que desempeña un estado cualitativo dado (junto con otros estados mentales) al mediar las conexiones entre entradas y salidas pudiera ser desempeñado por otro estado cualitativo, y si el primer estado cualitativo pudiera desempeñar un papel distinto, entonces no sería esencial para ese estado desempeñar dicho papel causal y no podría formar parte del significado, o sentido, de un término que lo designa rígidamente que el estado así designado es *el* estado que desempeña ese papel causal. Además, como tal término no podría ser eliminado de esta manera en favor de una descripción definida, no podría ocurrir dentro de la descripción definida que define funcionalmente el nombre de otro estado mental, suponiendo que el objetivo de tales definiciones funcionalistas sea eliminar la terminología mental en favor de una terminología física y tópico neutral.

Pero no hay nada en esto que implique que los estados cualitativos no puedan figurar entre los estados cuantificados en las descripciones definidas que definen otros tipos de estados mentales. Y parece que sería la cuantificación sobre tales estados, en vez de la referencia a los estados particulares de esta

[21] Véase Lewis 1972. Véase también Shoemaker 1981 o el ensayo 12 de Shoemaker 1984.

clase, lo que se necesitaría para definir otros estados mentales. Si la inversión del espectro es posible, no queremos convertir la ocurrencia de cualquier estado cualitativo particular en condición necesaria de ver (o de parecerle que ve) algo azul, pero sí queremos exigir que, en cualquier momento dado de la historia de una persona, haya uno u otro estado cualitativo que esté (en ese momento) normalmente implicado en su ver (o parecerle que ve) cosas azules. La especificación de los papeles de los estados cualitativos en ver cosas azules hará referencia indudablemente a las nociones de similitud y diferencia cualitativas; pero esto no representa dificultad alguna para un funcionalista si, como he sugerido, estas nociones pueden ellas mismas ser definidas funcionalmente.

Parecería haber, no obstante, algunos estados mentales (distintos de los estados cualitativos) que no pueden definirse funcionalmente en el sentido fuerte que aquí estamos considerando, a saber, en forma tal que no haya un uso esencial (no eliminable) de la terminología mental en el *definiens*. Pues considérense los estados que he llamado "creencias cualitativas", esto es, creencias sobre los estados cualitativos y, en particular, creencias de que uno (mismo) se encuentra en un estado cualitativo particular. Las creencias cualitativas pueden dividirse en dos grupos, aquellas en cuyo contenido proposicional hay una referencia a estados cualitativos particulares y aquellas en cuyo contenido proposicional hay una cuantificación sobre los estados cualitativos pero no hay referencia a estados cualitativos particulares. Hasta donde alcanzo a ver, las creencias cualitativas del segundo tipo no le presentan dificultades especiales al funcionalista; si otras clases de creencias pueden definirse funcionalmente, también éstas. Pero las creencias cualitativas del primer tipo sí parecen resistirse a una definición funcional. Considérese la creencia que yo expresaría si dijera: "Estoy en el estado de aparecérseme azul", utilizando la frase "estado de aparecérseme azul" para designar rígidamente un estado cualitativo particular. Si tratamos de caracterizar este estado de creer funcionalmente, es decir, en términos de sus relaciones con otros estados mentales y con entradas y salidas, parecería que, en nuestra caracterización, tendríamos que hacer referencia al estado cualitativo sobre el que versa la creencia; tendríamos que

decir que el estado de creer que a uno se le aparece azul es normalmente el resultado del estado de aparecerse azul a. Si esto es así, es imposible definir tales estados (creencias cualitativas del primer tipo) sin hacer uso esencial de términos mentales.

Pero esto no constituye un obstáculo para definir funcionalmente otras clases de estados mentales. Pues aunque queramos incluir en nuestras caracterizaciones funcionales de ciertas clases de estados mentales que dan origen a creencias cualitativas del primer tipo (es decir, aquéllas en cuyo contenido proposicional hay una referencia a estados cualitativos particulares), esto no requiere que hagamos una referencia identificatoria a creencias de este tipo en nuestras caracterizaciones funcionales; todo lo que se requiere es la cuantificación sobre tales creencias. Así, por ejemplo, podemos incluir en nuestra caracterización funcional del dolor que tener dolor normalmente provoca una creencia cualitativa de que uno está en cierto estado cualitativo específico, sin decir acerca de ningún estado cualitativo específico que sentir dolor tiende a dar origen a una creencia acerca de él. Y si es permisible cuantificar sobre estados cualitativos al ofrecer definiciones funcionales, no veo razón alguna para que no sea permisible también la cuantificación sobre creencias funcionales.

Volvamos ahora brevemente a mi argumento de la sección III contra la posibilidad de los casos de "*qualia* ausentes". En ese argumento indiqué que es característico de los dolores dar origen a una conciencia introspectiva de ellos como poseedores de caracteres cualitativos particulares, y así dar origen a "creencias cualitativas", y utilicé esto para argumentar que cualquier estado funcionalmente idéntico a un estado con carácter cualitativo (por ejemplo, un dolor) debe en sí mismo tener carácter cualitativo. Se planteó la objeción a este argumento de que, puesto que las creencias cualitativas, como los estados cualitativos, no pueden definirse en términos funcionales, no pueden entrar legítimamente en una explicación funcional de las "condiciones de identidad tipo" de otros estados mentales. Podemos ahora responder a esta objeción. Sin duda los dolores dan origen al tipo de creencias cualitativas que (según estoy concediendo) no pueden ser definidas funcionalmente, es decir, creencias de que uno está en un estado cualitativo específico. Pero también dan

origen a creencias de que uno siente dolor, y si (como parece suponer el "argumento de los *qualia* ausentes") el dolor es necesariamente un estado que posee carácter cualitativo, entonces la creencia de que uno tiene dolor conlleva supuestamente (por lo menos en el caso de una persona reflexiva) la creencia de que uno se encuentra en un estado que tiene tal o cual carácter cualitativo. Y mientras que esta última creencia es una creencia cualitativa, su contenido proposicional cuantifica sobre estados cualitativos en vez de hacer referencia a estados cualitativos particulares. No se ha ofrecido razón alguna de por qué no se deba considerar que las creencias cualitativas de este tipo son definibles en el nivel funcional. Y si son funcionalmente definibles, no hay razones para que la tendencia de otros estados a dar origen a tales creencias no deba formar parte de lo que constituye la identidad funcional de esos otros estados. Y esto es todo lo que requiere el argumento de la sección III.

VII

Durante las últimas décadas, gran parte de la controversia en filosofía de la mente ha entrañado una batalla entre dos conjuntos de intuiciones aparentemente en conflicto. Por un lado está la intuición de que los estados mentales se encuentran, de una u otra forma, vinculados lógica o conceptualmente con estados de cosas físicas, en particular, con las conductas que, se considera, los manifiestan. Esta intuición se ha expresado en una serie de posturas filosóficas diferentes: el conductismo lógico, las concepciones "criteriológicas" inspiradas por Wittgenstein y, más recientemente, los análisis funcionales o causales de los estados mentales (que suelen combinarse con alguna forma de materialismo o fisicalismo).[22] Por otro lado está la intuición de que las conexiones entre los estados mentales y la conducta son,

[22] Algunos defensores de las teorías causales o funcionales de la mente objetarían que se los coloque en este grupo. Pero otros han considerado claramente que sus explicaciones incorporan lo correcto en las concepciones conductistas y criteriológicas, o dan cuenta de las intuiciones que las hacen plausibles. Véanse, por ejemplo, Lewis 1972, p. 247, Armstrong 1968, p. 92, y Goldman 1970, p. 112.

en el fondo, contingentes; que conforme a las descripciones más "intrínsecas" de los estados mentales, es un hecho contingente que están relacionados como lo están con la conducta y con otros tipos de estados físicos. Y una expresión común de este punto de vista ha sido la afirmación de que la inversión del espectro y otros tipos de "inversión de *qualia*" son lógicamente posibles; pues decir que éstos son lógicamente posibles es, en apariencia, decir que cualquiera que sea el carácter intrínseco, interno, que tienen estos estados mentales, su "contenido cualitativo", éste carece de pertinencia lógica para la cuestión de que estén relacionados como lo están con sus causas corporales y sus manifestaciones conductuales. He concedido que esta idea tiene un elemento sustancial de verdad. Pues he permitido que la inversión del espectro sea una posibilidad y he permitido que esto implique que por lo menos algunos estados cualitativos (y creencias cualitativas) no puedan definirse funcionalmente. Pero creo que también la otra idea contiene un elemento sustancial de verdad. Pienso que donde la otra idea –la de que los estados mentales se encuentran vinculados "lógica" o "conceptualmente" con la conducta– resulta más plausible es en su aplicación a estados como el deseo y la creencia, y pienso que estos estados no poseen "carácter cualitativo" en el sentido que aquí nos ocupa, aunque en ocasiones pueden estar acompañados por estados cualitativos. Pero como ya he intentado mostrar, incluso los estados cualitativos pueden quedar dentro del marco de un análisis funcional, o causal, de los estados mentales. Aunque pueda pertenecer a la esencia de los estados cualitativos el que sean "inefables" en el sentido de que no puede decirse en términos generales, o en cualquier caso en términos generales que no incluyan nombres de estados cualitativos, lo que es para una persona hallarse en un estado cualitativo particular, esto no nos impide dar una explicación funcional de lo que es que un estado sea un estado cualitativo y de lo que son las condiciones de identidad de los estados cualitativos. Así, puede ser posible conciliar estas intuiciones, firmemente arraigadas y aparentemente en conflicto, sobre la contingencia o no contingencia de las relaciones existentes entre los estados mentales y el mundo físico.

Hay varias cuestiones que habría que investigar antes de poder afirmar que esta conciliación tentativa es exitosa. La explicación de la similitud y la diferencia cualitativas que he sugerido fue hecha para el caso de las experiencias perceptuales, y es preciso considerar si puede ser aplicada de manera plausible a sensaciones como los dolores. Lo que tal vez requiera su aplicación al caso del dolor es la aceptación de la idea de los dolores como impresiones sensoriales somáticas, es decir, impresiones (que no necesitan ser verídicas) de lesiones en el cuerpo y cosas semejantes.[23] Además, esta explicación de la similitud y la diferencia cualitativa está hecha para el caso en el cual las experiencias que se están comparando son experiencias de una y la misma persona, y es preciso considerar si provee de sentido, y del sentido correcto, a las comparaciones de experiencias intersubjetivas. Esto conllevaría, entre otras cosas, considerar si es posible que experiencias de personas distintas sean "coconscientes" en el sentido definido con anterioridad; y pienso que esto se reduce a la pregunta de si es posible que haya una "fusión" de personas en la forma contemplada en ciertas discusiones recientes sobre identidad personal, es decir, la unión de dos personas en una sola persona (o en un solo sujeto de conciencia) que luego recuerda, y es capaz de comparar, las experiencias que las personas tenían antes de la fusión. (Vale la pena observar que si la fusión es posible, entonces no es el caso, después de todo, que ninguna conducta posible revelaría si la experiencia de color de dos personas era la misma, o si sus espectros de color estaban invertidos el uno respecto del otro; pues si las personas se fundieran, la conducta de la persona resultante podría, supuestamente, resolver esta pregunta.) Pero éstas son cuestiones complejas, y no intentaré discutirlas aquí.

[*Traducción de Ana Isabel Stellino*]

[23] Esta idea ha sido en efecto propuesta por Armstrong y por Pitcher. Véanse Armstrong 1968, pp. 313 y ss., y Pitcher 1970.

BIBLIOGRAFÍA

Armstrong, D., 1968, *A Materialist Theory of the Mind*, Routledge and Kegan Paul, Londres.

Block, N.J. y J.A. Fodor, 1972, "What Psychological States are Not", *Philosophical Review*, vol. 81, pp. 159–181. [Incluido en esta antología, pp. 155–181.]

Chisholm, R., 1965, " 'Appear', 'Take', and 'Evident' ", en Swartz 1965.

———, 1957, *Perceiving: A Philosophical Study*, Cornell University Press, Ítaca.

Donagan, A., 1966, "Wittgenstein on Sensation", en Pitcher 1966, pp. 324–351.

Ginet, C., 1968, "How Words Mean Kinds of Sensations", *Philosophical Review*, vol. 77, pp. 3–24.

Goldman, A., 1970, *A Theory of Human Action*, Prentice-Hall, Englewood Cliffs.

Harrison, B., 1973, *Form and Content*, Blackwell, Oxford.

Kripke, S., 1972, "Naming and Necessity", en D. Davidson y G. Harman (comps.), 1972, *Semantics of Natural Language*, Reidel, Dordrecht, pp. 253–355. [Versión en castellano: *El nombrar y la necesidad*, 2a. ed. trad. M.M. Valdés, Instituto de Investigaciones Filosóficas-UNAM, México, 1995.]

———, 1971, "Identity and Necessity", en Munitz 1971, pp. 135–164. [Versión en castellano: *Identidad y necesidad*, trad. Margarita M. Valdés, Instituto de Investigaciones Filosóficas–UNAM, México, 1978 (Cuadernos de Crítica, 7).]

Lewis, D., 1972, "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy*, vol. 50, pp. 249–257.

Locke, D., 1968, *Myself and Others*, Clarendon Press, Oxford.

Munitz, K.M., 1971, *Identity and Individuation*, New York University Press, Nueva York.

Pitcher, G., 1970, "Pain Perception", *Philosophical Review*, vol. 79, pp. 368–393.

———(comp.), 1966, *Wittgenstein: The Philosophical Investigations*, Doubleday, Garden City.

Shoemaker, S., 1984, *Identity, Cause and Mind: Philosophical Essays*, Cambridge University Press, Cambridge.

———, 1982, "The Inverted Spectrum", *Journal of Philosophy*, vol. 79, pp. 357–381.

———, 1981, "Some Varieties of Functionalism", *Philosophical Topics*, vol. 12, no. 1, pp. 93–120.
———, 1975, "Phenomenal Similarity", *Crítica*, vol. 7, no. 20, pp. 3–37.
Swartz, R.J. (comp.), 1965, *Perceiving, Sensing and Knowing*, Anchor Books, Garden City, N.Y.

QUINEAR LOS *QUALIA**

DANIEL DENNETT

1. *En busca del azogue*

"*Qualia*" es un término poco conocido para algo que no podría resultarnos más conocido: las *maneras en que nos parecen las cosas*. Como sucede tan frecuentemente con la jerga filosófica, resulta más fácil ofrecer ejemplos que dar una definición del término. Mira un vaso de leche al atardecer: *la manera en que a ti te parece*, la cualidad visual particular, personal y subjetiva del vaso de leche es el *quale* de tu experiencia visual en ese momento. La *manera en que la leche te sabe en ese momento* es otro *quale*, el *quale* gustativo, y la *manera en que se oye* mientras lo tragas es un *quale* auditivo. Estas diversas "propiedades de la experiencia consciente" constituyen ejemplos magníficos de *qualia*. No hay

* La primera versión de este artículo se presentó en University College, Londres, en noviembre de 1978, y, con diversas revisiones, en una docena de otras universidades entre 1979 y 1980. Nunca fue publicado, pero tuvo una amplia circulación como Documento de Trabajo de Ciencias Cognoscitivas 7 de Tufts University, en diciembre de 1979. Se presentó una segunda versión en las universidades de Adelaida y Sydney en 1984 y, en 1985, en los coloquios de los departamentos de psicología de Harvard y de Brown, bajo el título "Properties of Conscious Experience". La segunda versión circuló como preimpreso en 1985, nuevamente bajo el título de "Quining Qualia". La presente versión, la cuarta, es una revisión sustancial, gracias a los útiles comentarios de muchas personas, entre quienes figuran Kathleen Akins, Ned Block, Alan Cowey, Sydney Shoemaker, Peter Bieri, William Lycan, Paul Churchland, Gilbert Harman, y los participantes de Villa Olmo.

nada, al parecer, que puedas conocer más íntimamente que tus propios *qualia*; aun cuando el universo entero fuera una vasta ilusión, una mera invención del demonio maligno de Descartes, aquello de lo que esa invención *estaría hecha* (para ti) serían los *qualia* de tus experiencias alucinatorias. Descartes afirmaba que dudaba de todo lo que podía ser dudado, pero jamás dudó que sus experiencias conscientes tuvieran *qualia*, las propiedades por las cuales las conocía o aprehendía.

El verbo "quinear"* es aún más esotérico. Proviene de *The Philosophical Lexicon* (Dennett 1987b), un diccionario satírico de epónimos: "quinear: v. Negar resueltamente la existencia o la importancia de algo real o significativo." A primera vista resulta difícil imaginar una tarea más quijotesca que tratar de convencer a la gente de que no existen propiedades como los *qualia*; de aquí el irónico título de este artículo. Pero no estoy bromeando.

Mi propósito es subversivo. Pretendo echar por tierra una idea que, de una u otra forma, es "obvia" para la mayoría de las personas: científicos, filósofos y legos. La mía es una empresa frustrantemente evasiva; apenas se hace retroceder a esa idea frente a un argumento, que "ella" reaparece, bajo un nuevo disfraz, aparentemente inocente de todos los cargos.

¿Cuál es la idea de *qualia* que intento extirpar? Todo lo real posee propiedades, y como no niego la realidad de la experiencia consciente, concedo que ésta posee propiedades. Concedo, además, que los estados de conciencia de cada persona poseen ciertas propiedades en virtud de las cuales esos estados tienen el contenido experiencial que tienen. Es decir, siempre que alguien experimenta algo como siendo de una manera y no de otra, esto es verdad en virtud de alguna propiedad de algo que sucede en esa persona en ese momento, pero estas propiedades son tan distintas de las que tradicionalmente se atribuyen a la conciencia que sería un gran error darle a cualquiera de ellas el nombre de los tan buscados *qualia*. Se supone que los *qualia* constituyen propiedades *especiales*, en un sentido difícil de definir. Mi afirmación —que sólo se esclarecerá a medida que

* El verbo "quinear" proviene del nombre del filósofo W.V.O. Quine. [N. de las comps.]

avancemos— es que la experiencia consciente no tiene ninguna propiedad que sea especial en *ninguno* de los sentidos en que se ha supuesto que los *qualia* son especiales.

La reacción común ante esta afirmación es el reconocimiento complaciente de que, mientras algunas personas pueden de hecho haber sucumbido a una confusión o a un fanatismo de una u otra especie, la alusión que uno mismo hace a una noción modesta e inocente de propiedades de la experiencia subjetiva está indudablemente a salvo. Es precisamente esa presunción de inocencia la que quiero echar por tierra. Quiero cambiar el peso de la prueba, de suerte que quien desee recurrir a propiedades subjetivas y privadas deba demostrar primero que, al proceder así, no está cometiendo un error. Esta condición de *culpable hasta que se demuestre lo contrario* no carece de precedentes ni resulta indefendible (en tanto nos limitemos a los conceptos). Hoy en día ningún biólogo se atrevería a suponer que sea enteramente correcto recurrir a algún concepto inocente de *élan vital*. Desde luego, uno *podría* usar el término para significar algo respetable; uno podría usar *élan vital* como el nombre que uno le da, por ejemplo, al ADN, pero resultaría una nomenclatura tonta, considerando la merecida suspicacia con que actualmente está teñido el término. Mi propósito es hacer que quienquiera que hable de *qualia* —o "sensaciones crudas" o "propiedades fenoménicas" o "propiedades subjetivas e intrínsecas" o "el carácter cualitativo" de la experiencia— se sienta igualmente incómodo al presuponer, como comúnmente se hace, que él, y todos los demás, saben de qué diablos están hablando.[1]

¿Qué son *exactamente* los *qualia*? Esta pregunta altisonante es desdeñada por un autor ("sólo medio en broma") recordando la legendaria respuesta que dio Louis Armstrong cuando le preguntaron qué era el *jazz*: "Si tienes que preguntar, nunca vas a saberlo" (Block 1978, p. 281). Esta divertida táctica ilustra a la perfección la presunción que constituye mi objetivo. Si tengo éxito en mi tarea, dicha maniobra, que actualmente es acepta-

[1] Una muestra representativa de la literatura más reciente sobre *qualia* incluiría a Block 1980; Shoemaker 1981, 1982; Davis 1982; White 1985; Armstrong y Malcolm 1984; Churchland 1985; y Conee 1985.

da en casi todos los círculos, parecerá tan extraña e intolerable como una alusión humorística a lo absurdo que resultaría una cosa viviente –¡una cosa viviente, fíjense!– que dudara de la existencia del *élan vital*.

Mi tesis, pues, no es sólo que los diversos conceptos técnicos o teóricos de los *qualia* son vagos o equívocos, sino que el concepto original, la noción "preteórica" de la que los primeros son supuestos refinamientos, es tan profundamente confusa que, incluso si nos obligáramos a salvar un "común denominador mínimo" de las propuestas de los teóricos, cualquier versión aceptable tendría que ser tan radicalmente distinta de las mal formadas nociones a las que se suele recurrir que sería tácticamente estúpido –por no decir *pickwickiano*– aferrarse al término. Es mucho mejor, desde el punto de vista táctico, declarar que, sencillamente, los *qualia* no existen en absoluto.[2]

Los argumentos rigurosos sólo funcionan con materiales bien definidos y, dado que mi objetivo es destruir nuestra fe en el concepto preteórico o "intuitivo", las herramientas correctas para mi tarea consisten en generadores de intuiciones, no en argumentos formales. Lo que sigue es una serie de quince generadores de intuiciones presentados en una secuencia que pretende sacar a la luz –y luego desechar– las intuiciones culpables. En la sección 2, utilizaré los dos primeros generadores de intuiciones para centrar la atención en la noción tradicional. Será tarea del resto del artículo convencerte de que estos dos generadores de intuiciones, pese a su eficacia, nos llevan por el camino equivocado y deben ser desechados. En la sección 3, los otros cuatro generadores de intuiciones crean y refinan una "paradoja" latente en la tradición. No se trata de una paradoja formal, sino sólo de un poderosísimo argumento que se opone

[2] La diferencia entre el "materialismo eliminativo" –del que mi postura sobre los *qualia* es un ejemplo– y un materialismo "reductivo" que asume la tarea de identificar la parte problemática en términos de la teoría materialista fundacional suele, pues, apreciarse mejor no tanto como una cuestión doctrinal sino como un asunto táctico: ¿cuál sería la forma más delicada o eficaz de iluminar a quienes están confundidos acerca de este caso? Véase mi discusión de las "fatigas" en la Introducción a *Brainstorms* (Dennett 1978a) y mi discusión previa sobre lo que deberían decir los iluminados acerca de la condición metafísica de beneficios [*sakes*] y las voces en *Content and Consciousness* (Dennett 1969, cap. 1).

a ciertas ideas que tienen un atractivo casi irresistible. En la sección 4, se exponen otros seis generadores de intuiciones con el fin de disipar la atracción que ejercen esas ideas, y la sección 5 remata la cuestión mostrando hasta qué punto resultan desafortunadas esas ideas cuando se las confronta con casos reales de experiencias anómalas. Esto dejará una especie de vacío, y en la sección 6 se utilizarán otros tres generadores de intuiciones para introducir y motivar algunas sustituciones adecuadas de las nociones desterradas.

2. *Las propiedades especiales de los* qualia

Generador de intuición 1: mirándote comer coliflor. Te veo saborear ansiosamente una porción de humeante coliflor, cuyo simple olor me provoca algunas náuseas, y me encuentro preguntándome cómo puedes disfrutar *ese sabor*, y entonces se me ocurre que probablemente la coliflor te sabe (¿te tiene que saber?) diferente. Ésta parece una hipótesis plausible, sobre todo porque sé que el mismo alimento a veces me sabe distinto en diversas ocasiones. Por ejemplo, el primer sorbo del jugo de naranja del desayuno sabe mucho más dulce que el segundo si entre uno y otro interpongo un bocado de panqué en jarabe de maple; pero después de uno o dos tragos de café, el jugo de naranja vuelve a saber (¿aproximada o exactamente?) como el primer sorbo. Ciertamente, queremos decir (o pensar en) tales cosas y, ciertamente, no cometemos un error garrafal cuando lo hacemos, por lo tanto... no cabe duda de que es perfectamente correcto hablar de *la forma en que le sabe el jugo a Dennett en el momento* t, y preguntar si es exactamente igual o diferente de *la forma en que le sabe el jugo a Dennett en el momento* t′ o *la forma en que le sabe el jugo a Jones en el momento* t.

Esta "conclusión" parece inocente, pero aquí mismo hemos ya cometido el gran error. El paso final presume que puedo aislar los *qualia* de todas las otras cosas que están sucediendo, por lo menos en principio o por mor del argumento. Lo que cuenta como *la forma en que le sabe el jugo a x* puede distinguirse, según se supone, de lo que constituye un mero acompañamiento, una causa contribuyente o un producto secundario de esa forma "central". Uno imagina oscuramente que toma tales casos y los

despoja poco a poco hasta llegar a lo esencial, dejando su residuo común, la manera en que las cosas se ven, se oyen, se sienten, saben o huelen para diversos individuos en momentos distintos, independientemente de la manera en que esos individuos sean estimulados o afectados no perceptualmente, e independientemente de cómo ellos se dispongan luego a comportarse o a creer. El error no consiste en suponer que, en la práctica, podemos realizar alguna vez o siempre este acto de purificación con certidumbre, sino en el error más fundamental de suponer que existe una propiedad residual semejante que pueda tomarse en serio, por más inseguros que resulten nuestros intentos reales de aislar las instancias.

Los ejemplos que nos seducen son abundantes en todas las modalidades. No puedo imaginar, nunca sabré ni podré saberlo jamás, al parecer, de qué manera sonaba Bach a los oídos de Glenn Gould. (Apenas recuerdo de qué manera sonaba Bach a mis oídos cuando yo era niño.) Y no puedo saber, al parecer, cómo es ser un murciélago (Nagel 1974)[3] o si tu ves lo que yo veo, en cuanto a colores, cuando ambos miramos un cielo claro y "azul". Los casos caseros nos convencen de la realidad de esas propiedades especiales –los sabores, aspectos visuales, aromas, sonidos subjetivos– que luego aislamos aparentemente para su definición mediante esta destilación filosófica.

Resulta difícil identificar el carácter especial de estas propiedades, pero se lo puede ver en acción en el *generador de intuición 2: la máquina catadora de vinos*. ¿Podrían los Hermanos Gallo reemplazar a sus catadores de vinos humanos con una máquina? Un "sistema computacional experto" en control de calidad y clasificación se encuentra probablemente dentro de los límites de la tecnología existente. Hoy en día sabemos lo suficiente sobre la química pertinente para construir transductores que reemplacen las papilas gustativas y los órganos olfativos (una visión de color fina sería quizá un asunto más problemático), y podemos imaginarnos utilizando las salidas de tales transductores como materia prima –de hecho, los "datos sensoriales"– para realizar complejas evaluaciones, descripciones y clasificaciones. Viértase la muestra en el embudo y, al

[3] Incluido en esta antología, pp. 45-63.

cabo de unos minutos u horas, el sistema imprimirá el examen químico, junto con un comentario: "un exuberante y suave Pinot, aunque carente de fuerza", u otras palabras semejantes. Tal máquina bien podría funcionar mejor que los catadores de vinos humanos en todas las pruebas razonables de precisión y consistencia que puedan idear los productores de vino,[4] pero *indudablemente*, no importa cuán "sensible" y "discernidor" se vuelva ese sistema, nunca tendrá ni disfrutará lo que *nosotros* tenemos y disfrutamos cuando catamos un vino: ¡los *qualia* de la experiencia consciente! Cualesquiera que sean las propiedades informacionales, disposicionales o funcionales que posean sus estados internos, ninguna será especial en la forma en que lo son los *qualia*. Si tú compartes esa intuición, crees que existen *qualia* en el sentido que pretendo demoler.

¿Qué hay de especial acerca de los *qualia*? Los análisis tradicionales sugieren que son ciertas fascinantes propiedades de segundo orden de esas propiedades. En primer lugar, como *nadie puede decirle* a otro, por elocuente que sea uno y por cooperativo e imaginativo que resulte su interlocutor, la manera exacta en que uno está actualmente viendo, saboreando, oliendo y demás, los *qualia* son *inefables* —de hecho, son casos paradigmáticos de cosas inefables. De acuerdo con la tradición, por lo menos parte de la razón de que los *qualia* sean inefables radica en que se trata de propiedades intrínsecas, lo que parece implicar, *inter alia*, que en cierto modo son atómicas e inanalizables. Puesto que son "simples" u "homogéneos", no hay nada en que basarse cuando intentamos describir tal propiedad a alguien que no esté familiarizado con la instancia particular en cuestión.

Además, las comparaciones verbales no son los únicos cotejos que se descartan. *Cualquier* prueba objetiva, fisiológica o "meramente conductual" —como las que pasa el imaginario sistema de catación de vinos— necesariamente erraría el blanco (según puede argumentarse plausiblemente), así que todas las comparaciones interpersonales de estas maneras de aparecerse resultan (aparentemente) sistemáticamente imposibles. En

[4] La verosimilitud de esta concesión depende menos de una alta consideración hacia la tecnología que de un escepticismo adecuado hacia las capacidades humanas, documentadas hoy en un fascinante estudio de Lehrer (1983).

otras palabras, los *qualia* son propiedades esencialmente *privadas*. Y, por último, como son propiedades de *mis experiencias* (no son hígado picado, ni propiedades de, digamos, el torrente sanguíneo de mi cerebro, ¿o no has estado prestando atención?), los *qualia* son en esencia directamente accesibles a la conciencia de su experimentador (sea lo que fuere que esto signifique), o constituyen propiedades de la experiencia propia con la cual uno está íntima o directamente familiarizado (sea lo que fuere que esto signifique), o "cualidades fenomenológicas inmediatas" (Block 1978) (sea lo que fuere que esto signifique). Son, al fin y al cabo, las propiedades precisas cuya apreciación nos permite identificar nuestros estados conscientes. Así, para resumir la tradición, se supone que los *qualia* constituyen propiedades de los estados mentales de un sujeto que son:

1) inefables

2) intrínsecas

3) privadas

4) directa o inmediatamente aprehensibles en la conciencia.

De este modo se introducen los *qualia* en la escena filosófica. Aparentemente, han sido propiedades muy significativas para algunos teóricos porque parecen interponer un obstáculo inevitable e insuperable para el funcionalismo o, en términos más amplios, para el materialismo o, en términos aún más amplios, para cualquier punto de vista o tratamiento del mundo que sea puramente objetivo y de "tercera persona" (Nagel 1986). Los teóricos de la postura contraria han echado abajo paciente e ingeniosamente todos los argumentos, y han dicho una gran cantidad de cosas correctas, pero han cometido un error táctico, según sostengo, al expresar de una u otra forma lo siguiente: "Nosotros, los teóricos, podemos habérnoslas con *los qualia* de los que hablan sin problemas; les mostraremos que sólo se equivocan ligeramente con respecto a su naturaleza." Lo que deberían haber dicho es: "¿Cuáles *qualia*?"

A algunos teóricos, mi cuestionamiento les parece absurdo o equivocado, porque, para comenzar, creen tener una noción mucho más débil y, por ende, menos vulnerable, de los *qua-*

lia. Piensan que estoy construyendo y derribando a un hombre de paja y, de hecho, preguntan: "¿Quién dijo que los *qualia* son los modos inefables, intrínsecos, privados y directamente aprehensibles en que las cosas le parecen a uno?" Como la esencia cuádruple de los *qualia* que he sugerido puede resultarles tendenciosa a muchos lectores, quizá sea iluminador considerar, brevemente, una alternativa que, en apariencia, es más moderada: los *qualia* son simplemente "los rasgos cualitativos o fenoménicos de la[s] experiencia[s] sensorial[es], cuya posesión hace que las experiencias se parezcan o difieran entre sí, cualitativamente, de la manera en que lo hacen" (Shoemaker 1982, p. 367). Seguramente, no pretendo negar *esos* rasgos.

Respondo: todo depende de lo que signifique "cualitativos o fenoménicos". Shoemaker contrasta a la similitud y la diferencia *cualitativas* con la similitud y la diferencia "intencionales", similitud y diferencia de las propiedades que una experiencia representa o "acerca de" las cuales es. Esto resulta bastante claro, pero, ¿qué hay, entonces, de lo "fenoménico"? Entre las propiedades no intencionales (¿y por ende cualitativas?) de mis estados visuales figuran sus propiedades fisiológicas. ¿Podrían ser estas propiedades mismas los *qualia* de los que habla Shoemaker? Se supone que es evidente, según entiendo, que estas clases de rasgos están descartados, porque no son "accesibles a la introspección" (Shoemaker, comunicación personal). Éstos son quizá rasgos de mi estado visual, pero no de mi *experiencia* visual. No son propiedades *fenoménicas*.

Pero entonces, otra similitud no intencional que comparten algunos de mis estados visuales es que tienden a hacerme pensar en ir a dormir. Creo que este rasgo suyo *sí* es accesible a la introspección, en cualquier interpretación común preteórica. ¿Es ésta una propiedad fenoménica o no? El término "fenoménico" no significa para mí nada evidente ni no tendencioso, y sospechosamente parece dirigirse de vuelta a las formas inefables, privadas y directamente aprehensibles en que las cosas le parecen a uno.[5]

[5] Shoemaker (1984, p. 356) parece estar acercándose renuentemente a la aceptación de esta conclusión: "Así, a menos que encontremos ciertas bases que nos permitan negar la posibilidad de la clase de situación prevista [...], parece que debemos elegir entre rechazar la explicación funcionalista de la

De hecho, sospecho que muchos están poco dispuestos a tomar en serio mi cuestionamiento radical, en gran medida porque desean profundamente que los *qualia* sean reconocidos. Para numerosas personas, éstos parecen ser la última trinchera defensiva de la interioridad y la inasibilidad de nuestra mente, un baluarte contra el mecanismo invasor. Están seguros de que debe haber *algún* camino firme que vaya desde los casos caseros hasta la formidable categoría de los filósofos, pues en caso contrario la ciencia irrumpiría en su último bastión de especialidad.

Esta condición especial de las presuntas propiedades posee una larga y distinguida tradición. Creo que fue Einstein quien nos advirtió una vez que la ciencia no podía darnos el *sabor* de la sopa. ¿Podría haberse equivocado ese hombre tan sabio? Sí, si interpretamos sus palabras como un intento por recordarnos los *qualia* que se ocultan eternamente de la ciencia objetiva dentro del santuario interno subjetivo de nuestras mentes. No existe nada semejante. Otro hombre sabio así lo afirmó: Wittgenstein (1958, esp. las pp. 91-100). De hecho, lo que dijo fue: "La cosa en la caja no pertenece en absoluto al juego del lenguaje; ni siquiera como un *algo*; pues la caja podría incluso estar vacía. —No, se puede 'cortar por lo sano' por la cosa que hay en la caja; se neutraliza, sea lo que fuere" (p. 100); y pasaba entonces a compensar sus apuestas diciendo: "¡No es un *algo*, pero tampoco una *nada*! La conclusión fue sólo que una nada podría servir tan bien como un algo sobre el que nada puede decirse" (p. 102). Tanto las observaciones de Einstein como las de Wittgenstein pueden someterse a una interminable exégesis; sin embargo, en vez de emprender la tarea de arbitrar esta Guerra de Titanes, prefiero asumir lo que bien puede ser una postura más radical que la de Wittgenstein.[6] Los *qualia* ni siquiera son "algo sobre

similitud cualitativa y rechazar la concepción estándar de *qualia*. Preferiría no tener que elegir, pero si me obligan a ello, rechazaría la concepción estándar de *qualia*."

[6] Shoemaker (1982) atribuye a Wittgenstein una concepción (reconociendo que "no queda perfectamente claro" que esto sea lo que en realidad Wittgenstein sostenía) que se aproxima mucho al punto de vista que defiendo aquí. Sin embargo, para Shoemaker, "parecería a primera vista que Wittgenstein estaba equivocado" (p. 360), afirmación que Shoemaker sustenta con un experimento

lo que nada puede ser dicho"; "*qualia*" es un término filosófico que no fomenta[7] más que confusión, y a final de cuentas no refiere en absoluto a propiedad ni rasgo alguno.

3. *Recuperación de la paradoja tradicional*

Los *qualia* no siempre han sido bien aceptados por los filósofos. Aunque muchos han pensado, junto con Descartes y Locke, que tenía sentido hablar de propiedades privadas e inefables de la mente, otros han argumentado que esto es un perfecto sinsentido, pese a la naturalidad con que surge de los labios. Vale la pena recordar la manera en que los *qualia* fueron supuestamente rehabilitados como propiedades que debían tomarse en serio después de los ataques que les infligieron Wittgenstein y los verificacionistas tachándolos de pseudohipótesis. La versión original del *generador de intuición 3: el espectro invertido* (Locke 1690; véase la edición de 1959) consiste en una especulación sobre dos personas: ¿cómo sé que tú y yo vemos el mismo color subjetivo cuando miramos algo? Como ambos aprendimos las palabras de los colores viendo objetos con colores públicos que se nos mostraban, nuestra conducta verbal coincidirá *aun cuando experimentemos colores subjetivos enteramente distintos*. La intuición de que esta hipótesis resulta sistemáticamente inconfirmable (y no desconfirmable, desde luego) siempre ha sido muy sólida, pero algunas personas se han sentido siempre tentadas a pensar que la tecnología podría (en principio) cerrar el hueco.

mental que está lejos de ser a primera vista, y que analiza erróneamente si el presente artículo está en lo correcto. (No hay una buena razón, al contrario de lo que sostiene Shoemaker, para creer que la *experiencia* de su sujeto sea sistemáticamente diferente de lo que era antes de la inversión.) Smart (1959) expresa una aprobación recelosa y parcial hacia la línea dura de Wittgenstein, pero no se abre camino claramente hacia un eliminativismo tan intransigente como el que sostengo aquí.

[7] En 1979 leí una versión anterior de este artículo en Oxford, con un comentario de John Foster, quien defendía los *qualia* con todas sus fuerzas, sosteniendo que "los *qualia* no deben ser quineados sino fomentados [*fostered*]". La simetría exige, desde luego, la siguiente definición para la última edición de *The Philosophical Lexicon* (Dennett 1987b): "*foster* [fomentar y de Foster], v. Afirmar resueltamente la existencia o la importancia de algo quimérico o insignificante."

Supongamos, en el *generador de intuición 4: la máquina de estimulación cerebral*, que existiera un aparato neurocientífico que se ajusta en tu cabeza y alimenta mi cerebro con tu experiencia visual (como en la película, *Brainstorm*, que no debe confundirse con el libro, *Brainstorms*). Con los ojos cerrados, doy un informe preciso acerca de todo lo que estás viendo, salvo por el hecho de que me maravilla que el cielo sea amarillo, la hierba roja y así con otras cosas. ¿No confirmaría esto, empíricamente, que nuestros *qualia* eran diferentes? Pero supongamos que el técnico entonces quita la clavija del cable conector, la invierte 180 grados y la vuelve a insertar en el enchufe. Ahora reporto que el cielo es azul, la hierba verde, y así sucesivamente. ¿Cuál es la orientación "correcta" de la clavija? El diseño y la construcción del mecanismo exigiría que su "fidelidad" se ajustara o calibrara de acuerdo con la normalización de los informes de ambos sujetos, de modo que habríamos regresado al punto de partida evidencial. La moraleja de este generador de intuiciones es que no es posible ninguna comparación intersubjetiva de los *qualia*, ni siquiera con una tecnología perfecta.

Así estaban las cosas hasta que alguien inventó la versión, supuestamente mejorada, del experimento mental: el espectro invertido *intra*personal. La idea parece habérseles ocurrido a varias personas en forma independiente (Gert 1965; Putnam 1965; Taylor 1966; Shoemaker 1969, 1975 [incluido en esta antología, pp. 183–212]; Lycan 1973). Probablemente Block y Fodor (1972 [incluido en esta antología, pp. 155–181]) la tienen en mente cuando expresan: "Nos parece que los contraargumentos verificacionistas estándares que se oponen a la idea de que la hipótesis del 'espectro invertido' es conceptualmente incoherente son poco convincentes" (p. 172) [p. 169–170]. En esta versión, *el generador de intuición 5: la jugarreta neuroquirúrgica*, las experiencias que se comparan están todas en una mente. Te despiertas una mañana para encontrarte con que la hierba se ha vuelto roja, el cielo amarillo, y así sucesivamente. Nadie más advierte ninguna anomalía con los colores del mundo, de modo que el problema debe estar en ti. Tienes derecho, al parecer, de concluir que has sufrido una inversión de los *qualia* visuales de los colores (y más adelante descubrimos, si se quiere, justo

cómo los neurofisiólogos malignos alteraron tus neuronas para lograr esto).

Aquí parece, al principio –y de hecho durante bastante rato–, que los *qualia* son, después de todo, propiedades aceptables, porque las proposiciones acerca de ellos pueden afirmarse con justificación, verificarse en forma empírica y hasta explicarse. Al fin y al cabo, en el caso imaginado, podemos contar una historia en la cual confirmamos una explicación neurofisiológica detallada de la etiología precisa del cambio drástico que has sufrido. Resulta tentador, pues, suponer que la evidencia neurofisiológica, incorporada a una teoría sólida y ramificada, tendría toda la capacidad de solución que pudiera ser necesaria para resolver, para determinar si los *qualia* de una persona han sido realmente alterados.

Pero esto es un error. Se requerirá una exploración paciente para poner al descubierto el error en toda su profundidad, pero se puede llegar rápidamente a la conclusión –si no asegurarla– con la ayuda del *generador de intuición 6: la neurocirugía alternativa*. Hay (por lo menos) dos formas diferentes en que el neurocirujano perverso podría crear el efecto de inversión descrito en el generador de intuición 5:

I. Invertir uno de los "primeros" canales productores de *qualia*, por ejemplo, en el nervio óptico, de modo que todos los sucesos neuronales pertinentes que van "corriente abajo" sean los "opuestos" de sus valores normales originales. *Ex* hipótesis, esto invierte tus *qualia*.

II. Dejar intactos todos esos primeros canales e invertir sólo algunos vínculos de acceso a la memoria, aquellos que logran tu comparación tácita (e incluso inconsciente) entre los tonos de hoy con los del pasado. *Ex* hipótesis, esto *no* invierte tus *qualia* en modo alguno, sino sólo las disposiciones que dependen de la memoria para reaccionar ante ellos.

Cuando despiertas y encuentras que tu mundo visual es sumamente anómalo, deberías exclamar: "¡Caray! ¡*Algo* ha ocurrido! He sufrido una inversión de mis *qualia* o de las reacciones ante ellos que dependen de mi memoria. ¡Me pregunto cuál de las dos!"

La suposición general fue que el experimento mental del espectro invertido intrapersonal constituía una mejora, ya que trasladaba la comparación requerida a la cabeza de un sujeto. Pero ahora podemos ver que esto es una ilusión, pues el vínculo con las experiencias anteriores, el vínculo establecido a través de la memoria, es análogo al cable imaginario que podría relacionar a dos sujetos en la versión original.

Esta cuestión es comúnmente —cabría decir "tradicionalmente"— pasada por alto por los constructores de los experimentos mentales de "espectros invertidos intrasubjetivos", quienes suponen que el hecho de que el sujeto *note la diferencia* —sin duda, una vívida experiencia de descubrimiento por parte del sujeto— tendría que ser una instancia de reconocer (¿directamente? ¿incorregiblemente?) tal diferencia como *una alteración de los qualia*. Pero como lo muestra mi ejemplo, podríamos llegar al mismo efecto desconcertante en un sujeto sin manipular sus *qualia* en lo absoluto. Como, *ex* hipótesis, las dos invasiones quirúrgicas distintas pueden producir exactamente los mismos efectos introspectivos, mientras que sólo una de ellas invierte los *qualia*, no hay nada en la experiencia del sujeto que favorezca una hipótesis sobre la otra. De modo que, a menos que busque ayuda externa, el estado de sus propios *qualia* debe resultarle tan incognoscible como el estado de los *qualia* de cualquier otra persona: ¡de ninguna manera el acceso privilegiado o la familiaridad inmediata o la aprehensión directa de las cuales los partidarios de los *qualia* suponían que gozaban los "rasgos fenoménicos"!

El resultado de esta serie de experimentos mentales constituye una intensificación del argumento "verificacionista" contra los *qualia*. *Si* existen *qualia*, éstos son aún menos accesibles a nuestra comprensión de lo que habíamos creído. No sólo son imposibles las comparaciones intersubjetivas clásicas (como lo demuestra la máquina de *Brainstorm*), sino que tampoco podemos saber en nuestro propio caso si nuestros *qualia* han sido invertidos, por lo menos no por introspección. Llegado a este punto, es sin duda tentador —especialmente para quienes no son filósofos— decidir que este resultado paradójico debe ser el producto de alguna especie de análisis filosófico equivocado, el tipo de cosa que bien puede suceder si se toma una noción

preteórica perfectamente buena —nuestra noción cotidiana de *qualia*— y se la fuerza ilícitamente más allá de sus límites. Los filósofos han creado un embrollo; dejemos que lo desenreden; mientras tanto, los demás podemos volver a nuestro trabajo, confiando como siempre en nuestro conocimiento sobrio no metafísico de los *qualia*.

Superar esta ubicua tentación es la tarea de la sección siguiente, la cual buscará establecer la insalvable incoherencia de las intuiciones que conducen a la paradoja examinando más detenidamente sus fuentes y su motivación.

4. *Errores que se cometen con respecto a los* qualia

La idea de que las personas podrían estar equivocadas con respecto a sus propios *qualia* yace en la médula de la confusión presente y debe ser explorada con mayor detalle y recurriendo a ejemplos un poco más realistas para entender el delicado papel que desempeña.

Generador de intuición 7: Chase y Sanborn. Había una vez dos catadores de café, el señor Chase y el señor Sanborn, quienes trabajaban para Maxwell House.[8] Junto con otros seis catadores, su trabajo consistía en cerciorarse de que el sabor del café de Maxwell House se mantuviera constante año con año. Un día, Chase, que llevaba ya seis años trabajando para Maxwell House, le confesó a Sanborn:

> Odio admitirlo, pero ya no disfruto este trabajo. Cuando llegué a Maxwell House hace seis años, pensé que su café era el más sabroso del mundo. Me enorgullecía compartir la responsabilidad de preservar ese sabor año tras año. Y hemos realizado bien nuestra labor; el café sabe exactamente igual hoy que cuando llegué. Pero fíjate, ¡ya no me agrada! Mis gustos han cambiado. Me he vuelto un bebedor de café más sofisticado. Ya no me gusta *ese sabor* en lo absoluto.

Sanborn acogió esta revelación con mucho interés. "Es gracioso que lo menciones", contestó, "pues me ha ocurrido algo muy parecido". Y prosiguió:

[8] Este ejemplo apareció impreso por primera vez en mis reflexiones sobre Smullyan, en *The Mind's I* (Hofstadter y Dennett 1981, pp. 427–428).

> Cuando llegué aquí, poco antes que tú, pensaba como tú que el café de Maxwell House era el de mejor sabor. Y ahora, como tú, no me importa realmente el café que estamos elaborando. Pero *mis* gustos no han cambiado; mis... *degustadores* son los que se modificaron. Es decir, creo que algo les ha ocurrido a mis papilas gustativas o a alguna otra parte de la maquinaria perceptual con la que analizo el sabor. El café de Maxwell House no me sabe igual que antes; si así fuera, me seguiría gustando, pues todavía pienso que *ese sabor* es el mejor sabor de café. Pero no estoy diciendo que no hayamos hecho bien nuestro trabajo. Ustedes, los otros catadores, están de acuerdo en que el sabor es el mismo, y debo admitir que, día tras día, yo tampoco puedo detectar ningún cambio. De modo que debe ser un problema exclusivamente mío. Supongo que ya no soy apto para este trabajo.

Chase y Sanborn se parecen por lo menos en un aspecto: a ambos solía gustarles el café de Maxwell House y ahora a ninguno le agrada. Sin embargo, afirman ser diferentes en otro aspecto. El café de Maxwell House a Chase le sabe igual que siempre, aunque no así a Sanborn. Pero, ¿podemos tomar sus declaraciones tal como las expresan? ¿Tenemos que hacerlo? ¿No podría uno de ellos o ambos estar equivocados? ¿No podrían sus dificultades ser fundamentalmente iguales y su aparente desacuerdo consistir más en una diferencia en la manera de expresarse que en una diferencia en un estado experiencial o psicológico? Puesto que ambos hacen afirmaciones que dependen de la confiabilidad de sus recuerdos, ¿hay alguna manera de verificar esta confiabilidad?

El motivo por el que introduje dos personajes en el ejemplo no es establecer una comparación interpersonal entre la manera en que el café le sabe a Chase y la manera en que le sabe a Sanborn, sino sólo mostrar, uno junto a otro, los dos polos entre los que pueden moverse los casos de las alteraciones de la experiencia intrapersonal. Tales casos de cambios en la experiencia intrapersonal, y la posibilidad de adaptarse a ellos, o la interferencia en ellos de la memoria, han sido discutidos frecuentemente en la literatura sobre los *qualia*, pero, en mi opinión, sin prestar suficiente atención a los detalles. Observemos primero a Chase. Si por el momento nos dejamos llevar por la

manera usual de hablar, parece inicialmente que tenemos las siguientes posibilidades:

(a) Los *qualia* de Chase asociados a la degustación del café han permanecido constantes, mientras que han variado sus actitudes reactivas hacia esos *qualia*, lo cual ha influido en sus cánones de juicio estético, etc. —lo que en su manera informal y espontánea de expresarse parece estar afirmando.

(b) Chase sencillamente se equivoca con respecto a la constancia de sus *qualia*; éstos se han alterado gradual e imperceptiblemente a través de los años, mientras que sus criterios gustativos no han cambiado, a pesar de su falsa creencia de que se han vuelto más sofisticados. Se encuentra en el estado que Sanborn afirma estar, sólo que carece del autoconocimiento que tiene este último.

(c) Chase está en un predicamento intermedio entre (a) y (b); sus *qualia* se han alterado un poco y sus estándares de juicio también se han modificado.

El caso de Sanborn parece estar sujeto a tres contrapartidas de las versiones del caso de Chase:

(a) Sanborn tiene razón; sus *qualia* se han alterado debido a algún tipo de trastorno de su maquinaria perceptual, pero, de hecho, sus estándares han permanecido constantes.

(b) Los estándares de Sanborn se alteraron sin que él se diera cuenta. Por ello está recordando erróneamente sus experiencias pasadas, en lo que podríamos llamar un efecto de nostalgia. Piensa en la conocida experiencia de reencontrarte con algún objeto de la infancia (un pupitre escolar, una casa en un árbol) y darte cuenta de que es mucho más pequeño de lo que recuerdas. Supuestamente, a medida que creciste, tu estándar interno acerca de lo que es grande, creció de una u otra forma junto contigo, pero los recuerdos (almacenados como fracciones o como múltiplos de ese estándar) no compensaron el cambio y, por ende, cuando consultas tu memoria, ésta te ofrece un juicio distorsionado. El recuerdo teñido de nostalgia que Sanborn tiene acerca del viejo y buen café de Maxwell House sufre una distorsión similar. (Hay, evidentemente, muchas maneras dis-

tintas en que puede implementarse este bosquejo impresionista del mecanismo de la memoria, y en la psicología cognoscitiva existe bastante trabajo experimental que sugiere cómo podrían comprobarse hipótesis diferentes sobre tal mecanismo.)

(c) Como en el caso anterior, el estado de Sanborn es cierta combinación de (a) y (b).

Creo que todos los que actualmente escriben sobre los *qualia* concordarán en que se dan todas estas posibilidades para Chase y Sanborn. No conozco a nadie que hoy en día se sienta tentado a defender la postura extrema de la infalibilidad o la incorregibilidad que establecería que la opción (a) es –y debe ser– la verdad en ambos casos, porque las personas simplemente no pueden equivocarse con respecto a estas cuestiones privadas y subjetivas.[9]

Sin embargo, como están a punto de surgir dificultades, conviene revisar esquemáticamente por qué el atractivo de la posición de la infalibilidad es sólo superficial, de modo que no pueda recuperar su previo encanto cuando las cosas se pongan difíciles. En primer lugar, a raíz de Wittgenstein (1958) y Malcolm (1956, 1959), hemos visto que una forma de lograr tal infalibilidad consiste en aceptar la completa evaporación del contenido (Dennett 1976): "¡Imagínate que alguien dijese: '¡Pero yo sé qué alto soy!', y a la vez llevara la mano como señal a su coronilla!" (Wittgenstein 1958, p. 96 [§279]). Al reducir la propia afirmación hasta que no quede nada que pueda ser correcto o incorrecto, uno puede lograr cierta vacua invencibilidad, pero esto no servirá en nuestro caso. Una de las cosas acerca de las cuales queremos que Chase esté en lo correcto (si lo está) es acerca de que no está en el mismo predicamento que Sanborn, de modo que, si ha de considerarse la afirmación como infalible, no pueda ser debido a que se niega a afirmar nada.

[9] Kripke (1982, p. 40) se aproxima a ello cuando pregunta retóricamente: "¿No sé, directamente y *con un buen grado de certidumbre*, que significo más [con la función que llamo 'más']?" (las cursivas son mías). Kripke no nos dice lo que está implicado en "un buen grado de certidumbre", pero supuestamente con esta afirmación pretende declarar su apoyo a lo que Millikan (1984) ataca bajo el nombre de "racionalismo de significado".

He visto que existe la poderosa tentación de responder a las afirmaciones que hago en este artículo más o menos como sigue: "Pero después de todo lo que se dice y se hace, sigue habiendo algo que sé de manera especial: *sé cómo son las cosas conmigo en este preciso instante.*" Sin embargo, si no se deriva absolutamente nada de este supuesto conocimiento –nada, por ejemplo, que arroje algo de luz sobre las distintas afirmaciones psicológicas que podrían ser verdaderas acerca de Chase o de Sanborn–, ¿qué caso tiene aseverar que uno lo posee? Quizá la gente sólo desea reafirmar su sentido de propiedad sobre sus propios estados conscientes.

La vertiente infalibilista de los *qualia* los trata como propiedades de la propia experiencia que, en principio, no pueden estar sujetas a error a la hora de descubrirlas, y ésta constituye una doctrina misteriosa (por lo menos, tan misteriosa como la infalibilidad papal) a menos que cambiemos un poco el énfasis y tratemos los *qualia* como *construcciones lógicas* a partir de los juicios de los sujetos sobre sus *qualia*: la experiencia de un sujeto tiene el *quale F* si y sólo si el sujeto juzga que su experiencia tiene el *quale F*. Podemos entonces tratar tales juicios como actos constitutivos que, de hecho, dan origen al *quale* por medio de la misma clase de licencia que usan los novelistas para determinar por decreto el color de cabello de sus personajes. No nos preguntamos cómo sabe Dostoievski que el cabello de Raskolnikov es castaño claro.

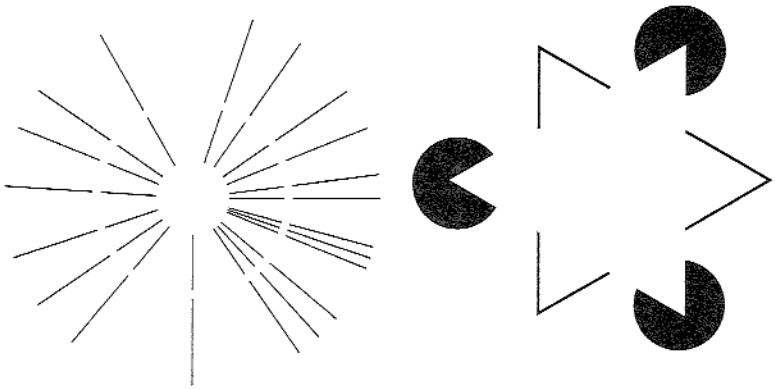
Se puede hacer un uso limitado de tales interpretaciones sobre los protocolos de los sujetos, según he argumentado (Dennett 1978a, 1979, esp. las pp. 109–110, 182), pero éstas no ayudarán a los defensores de los *qualia* en este punto. Las construcciones lógicas a partir de los juicios deben considerarse como afines a las ficciones de los teóricos, y los partidarios de los *qualia* quieren que la existencia de un *quale* particular en cualquier caso determinado sea un hecho empírico que se sostenga por sí mismo, no una útil ficción interpretativa de un teórico, pues de otro modo no aparecerá como un reto al funcionalismo, al materialismo o a la ciencia objetiva en tercera persona.

Parece bastante fácil, pues, imaginarse pruebas empíricas que tiendan a confirmar las distintas historias de Chase y San-

born, pero si pasar esas pruebas podría apoyar su autoridad (es decir, su confiabilidad), reprobarlas tendría que debilitarla. El precio que se paga por la posibilidad de confirmar empíricamente las afirmaciones propias consiste en la probabilidad externa de quedar desacreditado. Los partidarios de los *qualia* están dispuestos, hoy en día, a pagar ese precio, pero quizá sólo porque no han tomado en cuenta que el trato al que llegaron subvertirá el concepto que desean defender.

Consideremos de qué manera podríamos arrojar luz sobre la cuestión de dónde yace la verdad en los casos particulares de Chase y Sanborn, aun cuando quizá no seamos capaces de zanjar el asunto en forma definitiva. Es obvio que puede haber un apoyo objetivo revelador para una u otra versión extrema de sus historias. De este modo, si Chase es incapaz de volver a identificar cafés, tés y vinos durante cataduras a ciegas con intervalos de apenas unos minutos entre el primero y el segundo sorbos, su afirmación de que sabe que el café de Maxwell House le sabe exactamente igual que hace seis años quedará seriamente debilitada. Por el contrario, si sale airoso de las cataduras a ciegas y demuestra tener un considerable conocimiento sobre los cánones de los estilos de café (si tales existen), estará fundamentada su afirmación de que se ha vuelto un catador más sofisticado. La explotación de los principios estándar de la prueba inductiva –básicamente, el método de las diferencias de Mill– puede permitir un gran avance en la determinación del tipo de cambio que han sufrido Chase o Sanborn –un cambio próximo al extremo del procesamiento perceptual bruto del espectro o un cambio próximo al extremo del juicio reactivo final del espectro. Y como lo han observado Shoemaker (1982) y otros, las mediciones fisiológicas, interpretadas adecuadamente dentro de un marco teórico más amplio, también podrían inclinar la balanza hacia uno u otro extremo. Por ejemplo, se ha sostenido con frecuencia que el fenómeno, perfectamente estudiado, de los límites ilusorios inducidos (véase la figura) constituyen una ilusión particularmente "cognoscitiva", que depende de los procesos que van "de arriba abajo" y que, por lo tanto, se supone que están cercanos al extremo del espectro correspondiente al juicio reactivo, pero un trabajo experimental reciente (Von der Heydt *et al.* 1984) ha revelado que las neuro-

nas "detectoras de los bordes" *relativamente* bajas en los canales ópticos –en la zona 18 de la corteza visual– son tan responsivas a los bordes ilusorios como a los límites reales entre luz y oscuridad en la retina, lo cual sugiere (aunque no demuestra plenamente, ya que éste podría seguir siendo algún tipo de "efecto descendente") que los contornos ilusorios no son impuestos desde lo alto, sino que se generan en una etapa muy temprana del procesamiento visual. Cabe imaginar que se descubra en Sanborn una anomalía similarmente "temprana" en los canales que conducen de las papilas gustativas al juicio, por ejemplo, lo que tendería a confirmar su afirmación de que ha sufrido algún cambio en su maquinaria perceptual básica, y no en la del juicio.



Pero no sobreestimemos la capacidad de esa prueba empírica para llegar a una solución. El espacio existente en cada caso entre los dos polos representados por la posibilidad (a) y la posibilidad (b) sería ocupado por fenómenos que, de una u otra forma, son producto de dos factores en diversa proporción: toscamente, las disposiciones a generar o producir *qualia* y las disposiciones a reaccionar ante los *qualia* una vez que se han producido. (Así lo visualizaría nuestra imagen intuitiva de los *qualia*.) Se supone que los *qualia* sólo afectan nuestra acción o nuestro comportamiento por la mediación de los juicios acerca de ellos, de modo que cualquier prueba conductual, como una prueba de discriminación o de memoria, en cuanto toma los

actos basados en los juicios como sus datos primarios, sólo puede ofrecernos evidencia directa de la resultante de nuestros dos factores. En los casos extremos podemos contar con evidencia indirecta que sugiera que un factor ha variado mucho, mientras que el otro casi no, y podemos someter la hipótesis a una prueba adicional, checando la sensibilidad relativa del sujeto a variaciones en las condiciones que, supuestamente, alteran a los dos factores componentes. Pero no puede esperarse que esa prueba indirecta resuelva la cuestión cuando los efectos son relativamente pequeños, por ejemplo, cuando nuestras hipótesis rivales son la hipótesis predilecta de Chase, la (a), y la variante menor de que sus *qualia* han cambiado *un poco* y sus criterios se han modificado *menos de lo que él cree*. Esto será verdad aun cuando incluyamos en nuestros datos todos los efectos conductuales no intencionales o inconscientes, pues su peso será ambiguo. (Una respuesta latente más prolongada de Chase, ¿indicaría actualmente un proceso de "intento de renormalización de los *qualia*" o de "evaluación estética más amplia"?)

La limitada capacidad evidencial de la neurofisiología se advierte con particular claridad si imaginamos un caso de adaptación. Supongamos, en el *generador de intuición 8: la gradual recuperación postoperatoria*, que hemos "invertido quirúrgicamente" las conexiones de las papilas gustativas de Chase en la forma imaginaria estándar: después de la operación, el azúcar le sabe salada, la sal le sabe amarga, etc. Pero supongamos además –y esto es un supuesto tan realista como su negación– que Chase ha compensado posteriormente esa inversión, según lo demuestra su comportamiento. Ahora *dice* que la sustancia azucarada que le colocamos en la lengua es dulce, y ya no pide salsa de carne para su helado. Supongamos que la compensación es tan completa que en todas las pruebas conductuales y verbales su comportamiento es indistinguible del de los sujetos normales, y del que tenía él mismo antes de la operación.

Si todo el ajuste compensatorio interno se realizó durante una etapa temprana del proceso –intuitivamente "pre*qualia*"–, entonces sus *qualia* actuales se han restablecido para quedar exactamente como eran (con respecto a las fuentes externas de estimulación) antes de la cirugía. Si, por el contrario, parte o todo el ajuste compensatorio interno es "pos*qualia*", entonces

sus *qualia* no se han normalizado *aunque él crea que sí*. Pero los hechos fisiológicos, por sí solos, no arrojarán ninguna luz con respecto a dónde, dentro del curso del proceso fisiológico entre saborear y decir, debe trazarse la línea en la que los supuestos *qualia* aparecen como propiedades de esa fase del proceso. Los *qualia* son, desde luego, las propiedades "inmediatas o fenoménicas", pero esta descripción no nos servirá para localizar la fase correcta dentro del curso fisiológico, pues, repitiendo el generador de intuición 6, siempre habrá por lo menos dos formas posibles de interpretar la teoría neurofisiológica, independientemente de los resultados. Supongamos que nuestra teoría fisiológica nos dice (con tanto detalle como se quiera) que el efecto compensatorio se produjo en Chase mediante un *ajuste del proceso de acceso a la memoria*, el cual se requiere para que nuestra víctima compare los matices actuales con los pasados. Quedan *todavía* dos historias que podrían contarse:

(I) Los *qualia* actuales de Chase siguen siendo anormales, pero gracias a la revisión de su proceso de acceso a la memoria, de hecho ha ajustado sus recuerdos de cómo solían saber las cosas, de modo que ya no nota ninguna anomalía.

(II) El paso de la comparación con los recuerdos se produce justo antes que la fase de los *qualia* en la percepción del sabor; gracias a la revisión, ahora *produce* los mismos *qualia* del pasado ante los mismos estímulos.

En (I) los *qualia* contribuyen, de hecho, a la entrada en el comparador de recuerdos. En (II) forman parte de la salida del comparador de recuerdos. Éstas parecen ser dos hipótesis sustancialmente distintas, pero la evidencia fisiológica, por muy bien desarrollada que esté, no nos indicará en qué parte de la memoria debemos colocar a los *qualia*. La evidencia introspectiva de Chase tampoco resolverá el dilema entre (I) y (II), ya que, *ex* hipótesis, él no puede establecer una distinción confiable entre ambas historias. Recordemos que, en principio, fue para confirmar o desconfirmar la opinión de Chase que recurrimos a la evidencia neurofisiológica. No podemos utilizar al final su opinión para zanjar la cuestión entre nuestras teorías neurofisiológicas rivales. Chase puede pensar que piensa que

sus experiencias son las mismas que antes *porque* realmente lo son (y recuerda exactamente cómo solían ser), pero debe admitir que carece de recursos introspectivos para distinguir esa posibilidad de la alternativa (I), en la que piensa que las cosas son como solían ser *porque* sus recuerdos sobre la manera en que solían ser han sido distorsionados por sus nuevos hábitos compensatorios.

Ante la sistemática neutralidad de su materia de estudio, los fisiólogos pueden tener sus propias razones para preferir (I) sobre (II), o viceversa, ya que pueden haberse *apropiado* del término "*qualia*" para sus propios fines téoricos, con el propósito de denotar cierta familia de propiedades detectables que, les parece, desempeñan un papel importante en su teoría neurofisiológica del reconocimiento y de la memoria perceptuales. Chase o Sanborn podrían quejarse –junto con una buena cantidad de voceros filosóficos– de que estas propiedades que los neurofisiólogos optan por llamar "*qualia*" no son los *qualia* de los que ellos están hablando. La réplica de los científicos es la siguiente: "Si no podemos distinguir (I) de (II), ciertamente tampoco podemos apoyar ninguna de sus afirmaciones. Si quieren nuestro apoyo, deben renunciar a su concepto de *qualia*."

Lo sorprendente de esto no es sólo que los métodos empíricos no lograrían distinguir lo que parecen ser afirmaciones muy diferentes sobre los *qualia*, sino que no lograrían hacerlo *a pesar de constituir mejor evidencia que las convicciones introspectivas de los propios sujetos*. Pues los juicios del propio sujeto, al igual que las conductas o las acciones que los expresan, son el resultado de los dos factores que hemos postulado, y no pueden discernir las proporciones componentes mejor que las pruebas conductuales externas. De hecho, las convicciones "introspectivas" de un sujeto serán por lo general *peor* evidencia que lo que pueden captar observadores externos. Pues si nuestro sujeto es –como ocurre en la mayoría de los casos– un "sujeto ingenuo", ignorante de los datos estadísticos acerca de su propi caso o de casos similares, sus juicios francos e inmediatos son, en el nivel evidencial, iguales a los juicios perceptuales de cualquier observador ingenuo acerca de factores en el mundo externo. Los juicios intuitivos de Chase sobre la constancia de sus *qualia* no

son mejores, epistémicamente, que sus juicios intuitivos sobre, digamos, la constancia de la intensidad lumínica o la constancia de la temperatura ambiente —o la constancia de su propia temperatura corporal. Pasar a un estado corporal interno no modifica la intimidad de la relación epistémica de ninguna manera especial. ¿Tiene Chase fiebre o sólo se siente con fiebre? A menos que haya tomado medidas para calibrar y comparar su propio comportamiento, su opinión de que su aparato perceptor de la fiebre no ha sufrido ningún trastorno no es más que una corazonada. De modo semejante, Chase puede tener una opinión firme de hasta qué punto su aparato perceptor de sabores ha preservado su integridad, y hasta qué punto su juicio ha evolucionado por medio de la sofisticación, pero, mientras no se tengan los resultados del tipo de prueba laboriosa de tercera persona que acabamos de imaginar, sería un tonto si afirmara saber —especialmente saber de manera directa o inmediata— que el suyo era un claro caso de (a), un caso más cercano a (a) que a (b) o un caso cercano a (b).

Chase pisa terreno muy firme, epistémicamente, cuando reporta que *la relación* entre su actividad de sorber café y su actividad de juzgar ha cambiado. Recordemos que éste es el factor que Chase y Sanborn tienen en común: les solía gustar el café de Maxwell House, y ahora ya no. Sin embargo, a menos que Chase lleve a cabo en sí mismo los tipos de pruebas a las que otros podrían someterlo, sus convicciones sobre lo que ha permanecido constante (o casi constante) y lo que ha variado *deben ser meras adivinanzas*.

Pero entonces los *qualia* —suponiendo por el momento que sabemos de lo que estamos hablando— deben perder una de sus propiedades "esenciales" de segundo orden: lejos de ser propiedades directa o inmediatamente aprehensibles de nuestra experiencia, son propiedades cuyos cambios o constancias o bien están completamente fuera de nuestra comprensión, o bien son inferibles (en el mejor de los casos) de los exámenes desde "la tercera persona" de nuestros patrones de reacciones conductuales y fisiológicas (si Chase y Sanborn aceptan el sentido que los neurofisiólogos dan al término). Desde esta perspectiva, Chase y Sanborn no deberían ser considerados sujetos instrospectores capaces de una visión privilegiada de estas

propiedades, sino como autopsicólogos, teóricos cuyas convicciones sobre las propiedades de sus propios sistemas nerviosos están basadas no sólo en sus convicciones "inmediatas" o actuales acerca de sus experiencias, sino también en su apreciación del peso que tienen los sucesos del pasado reciente que recuerdan.

Hay, como veremos, buenas razones para que los neurofisiólogos y otros teóricos "objetivos desde la tercera persona" seleccionen esa clase de propiedades para estudiarlas. Pero no son *qualia*, por la sencilla razón de que la relación epistémica que uno tiene con ellas es exactamente igual a la relación epistémica que uno tiene con esas propiedades externas, pero detectables pronta, aunque faliblemente, como la temperatura ambiente o el peso. La idea de que uno debería consultar a un experto externo y someterse a complejas pruebas conductuales para confirmar qué *qualia* uno tuvo, nos desvía mucho indudablemente de nuestra idea original de los *qualia* como propiedades con las cuales tenemos una familiaridad particularmente íntima.

De modo que quizá hemos tomado el camino equivocado. La doctrina que nos condujo a este bochornoso resultado fue la que distinguía claramente a los *qualia* de sus efectos (normales) sobre las reacciones. Volvamos a considerar a Chase. Afirma que el café sabe "exactamente igual" que siempre, pero admite —mejor dicho, insiste en— que su reacción a ese sabor no es la que solía ser. Es decir, pretende ser capaz de divorciar su aprehensión (o recuerdo) del *quale* —el sabor, en el habla común— de sus distintas reacciones al sabor. Pero esta aprehensión o recuerdo es de suyo una reacción al presunto *quale*, de modo que Chase está haciendo algún juego de prestidigitación, sin duda inocentemente. Así, supongamos más bien que Chase había insistido en que precisamente *porque* su reacción era ahora diferente, el sabor había cambiado para él. (Cuando le contó a su esposa su versión original, ella le dijo: "¡No seas tonto! ¡Una vez que le añades el desagrado, cambias la experiencia!", y cuanto más reflexionaba en ello, más se convencía de que ella tenía razón.)

Generador de intuiciones 9: el bebedor de cerveza experimentado. Se dice comúnmente que la cerveza, por ejemplo, es un gusto adquirido; uno se entrena poco a poco para disfrutar su sabor,

o sólo llega a disfrutarlo. ¿Qué sabor? ¿El del primer trago? A nadie podría gustarle ese sabor, contestaría un bebedor de cerveza experimentado:

> La cerveza le sabe diferente al bebedor de cerveza experimentado. Si la cerveza siguiera sabiéndome como el primer sorbo que probé, ¡nunca habría vuelto a tomar cerveza! O para decirlo de otro modo, si mi primer sorbo de cerveza me hubiera sabido como me supo el más reciente, por principios de cuenta ¡nunca habría tenido que adquirir el gusto! Me habría encantado tanto el primer trago como el que acabo de disfrutar.

Si aceptamos este discurso, tendremos que admitir que la cerveza no es un gusto adquirido. Nadie llega a disfrutar *la manera en que le supo el primer sorbo*. Por el contrario, beber cerveza durante un tiempo prolongado hace que las personas experimenten un sabor que disfrutan, pero precisamente el hecho de que disfruten el sabor garantiza que no es el que experimentaron la primera vez.[10]

Pero esta conclusión, si se acepta, causa un tipo distinto de estragos en la concepción filosófica tradicional de los *qualia*. Pues si se admite que las actitudes o reacciones que uno tiene ante las experiencias son, de alguna manera y en algún grado, constitutivas de sus cualidades experienciales, de suerte que un cambio en la reactividad *equivalga a* o *garantice* un cambio en la propiedad, entonces esas propiedades, esos "rasgos cualitativos o fenoménicos", dejan de ser propiedades "intrínsecas" y, de hecho, se convierten en propiedades paradigmáticamente extrínsecas y relacionales.

Con frecuencia las propiedades que al principio "parecen intrínsecas" resultan ser, tras un análisis más detenido, relacionales. Bennett (1965) es el autor del *generador de intuiciones 10: el experimento mundial de eugenesia*. Nos llama la atención sobre la fenoltiourea, una sustancia que para tres cuartas partes de la humanidad tiene un sabor muy amargo, y que para el resto es

[10] Podemos salvar la afirmación tradicional pasando por alto los *qualia* supuestamente privados o subjetivos y hablando siempre de gustos públicos, como el gusto público del café de Maxwell House que, según concuerdan tanto Chase como Sanborn, ha permanecido constante. Se puede decir que los individuos adquieren un gusto por ese sabor público.

tan insabora como el agua. ¿Es amarga? Como la reactividad a la fenoltiourea se transmite genéticamente, podríamos volverla paradigmáticamente amarga realizando un experimento genético a gran escala: impedir que las personas a quienes les resulta insabora se reproduzcan y, en unas cuantas generaciones, la fenoltiourea sería tan amarga como lo más amargo del mundo. Pero también podríamos (en principio) llevar a cabo el experimento opuesto de "eugenesia" masiva, haciendo así que la fenoltiourea se volviera paradigmáticamente insabora —tanto como el agua— sin siquiera tocarla. Es evidente que el amargor o la falta de sabor público no constituye una propiedad intrínseca de la fenoltiourea, sino una propiedad relacional, puesto que la propiedad se altera mediante un cambio en la clase referencial de detectores normales.

Las versiones públicas de los "*qualia*" perceptuales *parecen* todas intrínsecas, a pesar de su carácter relacional. No son las únicas. Pensemos en el "valor sentido" de un dólar (o de cualquiera que sea la moneda de tu país). "¿Cuánto vale eso en dinero *verdadero*?", se dice que pregunta el turista estadounidense, esperando convertir cierto precio en moneda extranjera a la escala de "valor intrínseco" que él tiene en mente. Como lo afirma Elster (1985), "existe la tendencia a pasar por alto el carácter implícitamente relacional de ciertos predicados monádicos". Walzer (1985) señala que "podría parecer que un billete de diez dólares tiene vida propia como cosa de valor, pero, como sugiere Elster, su valor depende implícitamente de 'otras personas que están dispuestas a aceptar el dinero como pago de bienes' ". Sin embargo, incluso concediendo esto, todavía hay cierta tendencia a preservar algo subjetivo, el valor sentido, como una propiedad "intrínseca" de ese billete de diez dólares. Pero como ahora vemos, esas propiedades intrínsecas no pueden ser propiedades a las que un sujeto accede en forma privilegiada alguna.

¿Qué camino debe tomar Chase? ¿Debe seguir el consejo de su esposa y declarar que, como ya no puede soportar el café, ha dejado de saberle igual (solía saber bien y ahora sabe mal)? ¿O debe decir que, en cierto sentido, realmente sí sabe como siempre, o por lo menos relativamente igual, una vez que, desde luego, se sustrae el hecho de que ahora sabe muy mal?

Hemos llegado ya a la médula de mi argumentación. El hecho es que tenemos que preguntarle a Chase qué camino quiere seguir, y lo cierto es que tiene a su disposición dos opciones radicalmente distintas *si lo forzamos a elegir*. ¿Qué camino tomarías *tú*? ¿Qué concepto de *qualia* has tenido "siempre en el fondo de tus pensamientos", guiando tu imaginación cuando reflexionabas sobre las teorías? Si reconoces que la respuesta no es obvia, y sobre todo si objetas que esta elección forzada separa dos aspectos que habías supuesto que se hallaban unidos en su concepto preteórico, apoyarás mi aseveración de que no existe un fundamento firme en la "psicología popular" ordinaria para un concepto de *qualia*. *Normalmente* pensamos de manera confusa y potencialmente incoherente cuando reflexionamos acerca de las maneras en que las cosas nos parecen.

Cuando Chase piensa en "ese sabor", lo hace en forma equívoca o vaga. Se remonta a los recuerdos de experiencias pasadas, pero no necesita tratar –ni ser capaz– de decidir si está incluyendo algunas o todas sus reacciones o excluyéndolas de lo que pretende significar con "ese sabor". Su estado de entonces y su estado actual son diferentes –*eso* es algo que puede admitir con confianza–, pero carece de recursos "inmediatos" para realizar una distinción más fina, y no tiene ninguna necesidad de hacerlo.[11]

Esto sugiere que los *qualia* no son más esenciales al vocabulario profesional del fenomenólogo (o catador de café profesional) que al vocabulario del fisiólogo (Dennett 1978b). Para comprender esto, volvamos a considerar el ejemplo de mi desagrado por la coliflor. Imaginemos ahora, en el *generador de intuiciones 11: la cura de la coliflor*, que alguien me ofrece una píldora para curar mi rechazo a este alimento. Me promete que, cuando tome esta píldora, la coliflor me sabrá exactamente igual que siempre, pero ese sabor me gustará. "Espera", podría contestarle, "creo que acabas de contradecirte". Pero de todos modos

[11] "No estoy tan loco como para negar que mi sensación de rojo hoy es igual a mi sensación de rojo de ayer. Sólo digo que la similitud quizá consista únicamente en la fuerza fisiológica que subyace a la conciencia, la cual me lleva a afirmar que reconozco esta sensación como la misma a la primera, y por lo tanto no consiste en una comunidad de sensaciones." (Peirce 1931–1958, vol. V, p. 172, n. 2.)

tomo la píldora y ésta funciona. Me convierto repentinamente en un amante de la coliflor, pero si me preguntan cuál de los dos efectos posibles (el de tipo Chase o el de tipo Sanborn) ha tenido en mí esa píldora, estaré desconcertado y no encontraré nada *en mi experiencia* que arroje luz sobre el asunto. Desde luego, reconozco que el sabor es (en cierta medida) el mismo —la píldora, al fin y al cabo, no ha hecho que la coliflor tenga el sabor a pastel de chocolate—, pero al mismo tiempo mi experiencia es ahora tan diferente que me resisto a decir que la coliflor sabe como solía saber. En cualquier caso, no hay razón para que por fuerza suponga que mis experiencias con la coliflor poseen ciertas propiedades intrínsecas detrás, o además, de sus diversas propiedades disposicionales, causantes de reacciones.

"Pero en principio tiene que haber una respuesta correcta a la cuestión de cómo es, intrínsecamente, para ti ahora, aun cuando no puedas expresarlo con plena confianza." ¿Por qué? ¿Diría uno lo mismo de todas las otras propiedades de la experiencia? Consideremos el *generador de intuiciones 12: la inversión del campo visual creada por el uso de anteojos inversores,* un fenómeno que ha sido estudiado empíricamente durante años. (G.M. Stratton publicó la obra pionera en 1896, y J.J. Gibson e I. Kohler figuraron entre los principales investigadores; para una explicación introductoria, véase Gregory 1977.) Después de usar los anteojos inversores durante varios días, los sujetos se adaptan con un éxito sorprendente. Supongamos que los confrontamos con la siguiente pregunta: "¿Tu adaptación consiste en que has reinvertido tu campo visual o en que has trastocado por completo el resto de tu mente en un sinnúmero de compensaciones?" Si vacilan, ¿podemos insistir en que tiene que haber una respuesta correcta, aun cuando no puedan decir con plena confianza cuál es? Tal insistencia conduciría directamente a una nueva versión del viejo experimento de pensamiento del espectro invertido: "¿Cómo sé si algunas personas ven las cosas al revés (pero están perfectamente acostumbradas a ello), mientras que otras las ven al derecho?"

Sólo una concepción muy ingenua de la percepción visual podría sustentar la idea de que el campo visual de una persona incluye la propiedad de estar al derecho o al revés *independiente*

de sus disposiciones a reaccionar ante ellas –podríamos llamarla "el lado derecho intrínseco" (véase mi discusión sobre las propiedades de las "imágenes" procesadas por el robot SHAKEY en Dennett 1982). Así, no todas las propiedades de la experiencia consciente invitan o exigen tratarlas como "intrínsecas". ¿Hay algo distintivo de cierta subclase de propiedades (seguramente, la subclase "cualitativa o fenoménica") que nos obligue a tratarlas –al contrario del lado derecho subjetivo– como propiedades intrínsecas? Si no lo hay, tales propiedades no tienen ningún papel que desempeñar en las teorías fisiológicas de la experiencia ni en las teorías introspectivas.

Algunos podrían sentirse inclinados a argumentar de la siguiente manera: puedo imaginar definitivamente la experiencia de la "inversión del espectro" desde el interior; después de todo, de hecho he experimentado efectos temporales del mismo tipo, como el efecto del "desplazamiento del sabor" del jarabe de maple sobre el jugo de naranja. Lo que resulta imaginable, o real, es posible. Por tanto, la inversión o el desplazamiento del espectro (en todas sus modalidades sensoriales) es posible. Pero tales fenómenos *son* justamente la inversión o el desplazamiento de los *qualia*, o de las propiedades subjetivas intrínsecas. Por ende, deben existir los *qualia*: propiedades subjetivas intrínsecas.

Esto es falaz. Lo que uno imagina y lo que uno dice que imagina pueden ser dos cosas diferentes. Imaginar la inversión del campo visual, como lo experimentaron los sujetos de Stratton y Kohler, no equivale necesariamente a imaginar la inversión absoluta del campo visual (aun cuando esto sea "como lo sienten" los sujetos). De modo menos obvio, imaginar –tan vívidamente como se quiera– un caso de desplazamiento subjetivo en la percepción de los colores no es necesariamente imaginar el fenómeno al que los filósofos normalmente llaman: un espectro invertido o desplazado *de los qualia*. En la medida en que ese término conlleva las implicaciones problemáticas que aquí se exploran, no se puede apoyar su uso simplemente en lo vívido o la naturalidad de la posibilidad imaginada.

Si no existen propiedades como los *qualia*, ¿significa esto que la "inversión del espectro" es imposible? Sí y no. La inversión del

espectro según la discusión clásica es imposible, pero algo semejante resulta perfectamente posible, algo que se parece tanto a la "inversión de los *qualia*" como la inversión del campo visual se asemeja a la imposible inversión absoluta de la imagen visual que acabamos de desechar.

5. *Algunos casos reales desconcertantes*

No basta ocultar nuestras alianzas teóricas hasta el soleado día en que los filósofos completen la difícil tarea de purificar el concepto cotidiano de *qualia*. A menos que tomemos medidas diligentes para deshacernos de este concepto original y sustituirlo por mejores ideas, seguirá debilitando nuestra imaginación y distorsionando sistemáticamente nuestros intentos por comprender los fenómenos ya vistos.

Lo que descubrimos al observar, por ejemplo, los fenómenos reales de anomalías en la percepción del color, por ejemplo, confirma ampliamente nuestras sospechas sobre la inadecuación de la noción tradicional de *qualia*. Se han reportado algunas variedades de *acromatopsia cerebral* (una deficiencia cerebral en la visión de color), y aunque queda mucho sin resolverse en su análisis, hay pocas dudas de que los experimentos de pensamiento han subestimado o pasado por alto las posibilidades que existen para las colecciones de síntomas contraintuitivos, como lo muestran algunos breves extractos de los estudios de casos.

> Los objetos ubicados a la derecha del meridiano vertical parecían ser de color normal, mientras que aquellos a la izquierda se percibían sólo en tonos de gris, aunque sin distorsiones en la forma [...]. No era capaz de reconocer o nombrar ninguno de los colores en cualquier porción del campo izquierdo de cualquiera de sus ojos, ni siquiera los rojos, azules, verdes y amarillos brillantes. En cuanto una parte cualquiera del objeto coloreado cruzaba el meridiano vertical, podía reconocer el color instantáneamente y nombrarlo con precisión. (Damasio *et al.* 1980)

En un principio, parecería que este paciente puede ser descrito sin problemas como alguien que sufre una variación o una

pérdida de los *qualia* de color en el hemicampo visual izquierdo, pero aquí hay una dificultad de interpretación, puesta de manifiesto por otro caso:

> El paciente fracasaba en todas las tareas que le exigían aparcar el color visto con su nombre hablado. Así, el paciente no logró dar los nombres de los colores ni escoger el color en respuesta al nombre en cuestión. En contraste, tuvo éxito en todas las tareas en las cuales la vinculación era puramente verbal o puramente no verbal. Así, fue capaz de dar verbalmente los nombres de los colores que correspondían a los objetos nombrados y viceversa. Podía casar los colores vistos entre sí y con imágenes de objetos, y podía clasificar los colores sin error. (Geschwind y Fusillo 1966)

Este segundo paciente no se había percatado en lo absoluto de que sufriese alguna deficiencia. "Jamás respondía con un simple 'no sé' cuando se le pedía nombrar un color" (Geschwind y Fusillo 1966, p. 140). Hay un sorprendente contraste entre estos dos pacientes: la capacidad de ambos para nombrar los colores de las cosas que están por lo menos en parte de su campo visual es deficiente, pero mientras que el primero tiene plena conciencia de su deficiencia, el segundo no. ¿Acaso esta distinción representa toda la diferencia con respecto a los *qualia*? De ser así, ¿qué diablos debemos decir de este tercer paciente?

> Su otra gran queja era que "todo parecía negro o gris" y esto le causaba algunos problemas en su vida cotidiana [...]. Tenía una dificultad considerable en reconocer y nombrar los colores. Solía describir, por ejemplo, los objetos de color rojo vivo como rojos o negros, los objetos de color verde vivo como verdes, azules o negros, y los de color azul vivo como negros. La dificultad parecía ser perceptual, y las observaciones que hacía sugerían esto; por ejemplo, cuando se le mostraba un objeto de color rojo vivo decía: "un rojo ahumado y sucio, no tan rojo como normalmente uno ve el rojo". Describía los colores menos saturados o brillantes con términos como "gris", "blancuzco" o "negro", pero si se le pedía que adivinara el color, acertaba el cincuenta por ciento de las veces, y lograba un éxito notablemente menor con los azules y verdes que con los rojos. (Meadows 1974)

La conciencia que tenía este hombre acerca de su deficiencia resulta problemática, por decir lo menos. Contrasta muy marcadamente aún con otro caso más:

> Una mañana en noviembre de 1977, al despertar, notó que aunque podía ver los detalles de los objetos y las personas, los colores parecían "deslavados" e "irreales". No se quejaba de otra cosa [...]; su visión era buena, 20/20 en cada ojo [...]. La dificultad en la percepción de los colores persistía, y debía pedirle consejo a su esposo para escoger la ropa que iba a ponerse. Ocho semanas después advirtió que ya no podía reconocer las facciones de su esposo y de su hija [...]. [Así que] además de acromatopsia, la paciente sufría prosopagnosia, pero en otros aspectos su desempeño lingüístico y cognoscitivo no fue afectado. La paciente podía contar coherentemente lo que le había sucedido y tenía una notable comprensión de sus deficiencias. (Damasio *et al.* 1980)

Como lo observa Meadows: "Así algunos pacientes, se quejan de que su visión de los colores es deficiente, mientras que otros no se quejan en forma espontánea, pero muestran fuertes anormalidades en las pruebas."

¿Qué decir de estos casos? Cuando no hay queja que presente el paciente, pero éste muestra una deficiencia en la visión de colores, ¿es esto un indicio de que sus *qualia* no están afectados? ("Su capacidad de discriminación se encuentra profundamente deteriorada, pero, por fortuna para él, su vida interior permanece intacta por esta pérdida meramente pública.") Podríamos alinear los *qualia* de esta manera, pero de igual modo podríamos afirmar que el paciente sencillamente no se ha percatado de que sufre un, quizá, gradual deslave o inversión o fusión de sus *qualia*, que se manifiesta en su mal desempeño. ("Su vida interior perdió su complejidad y su variedad tan lentamente que nunca se dio cuenta de cuánto se había empobrecido.") ¿Qué pasaría si nuestra última paciente hubiera descrito su trastorno tal como lo hemos citado, pero se hubiera desempeñado normalmente en las pruebas? Una hipótesis sería que sus *qualia* se habían vuelto en efecto deslavados, tal como ella lo sugirió. Otra sería que, a la luz de su excelente desempeño en las pruebas de discriminación de los colores, sus *qualia* estaban bien; padecía alguna anomalía histérica o depresiva, una especie de hipocondría de

la visión en colores que la lleva a quejarse de una pérdida en la percepción de color. O quizá podría afirmarse que sus *qualia* permanecían intactos; su trastorno era puramente verbal: una comprensión anómala de las palabras que usa para describir su experiencia. (En la literatura se han reportado otros trastornos asombrosamente específicos relativos a las palabras de colores.)

El concepto tradicional nos lleva a pasar por alto posibilidades genuinas. Una vez que nos enteramos de la curiosa deficiencia de la que informan Geschwind y Fusillo (1966), por ejemplo, caemos en la cuenta de que a nuestro primer paciente nunca se lo examinó para ver si aún podía clasificar los colores que veía a la izquierda o pasar otras pruebas de daltonismo que no fueran verbales ni involucraran nombrar colores. Tales pruebas no son nada superfluas. Quizá las habría pasado; quizá, *a pesar de lo que dice*, sus *qualia* siguen tan intactos para el campo izquierdo como para el campo derecho, si tomamos la capacidad salir bien de esas pruebas como "criterio". Tal vez su problema sea "puramente verbal". Si tu reacción a esta hipótesis es que es imposible, esto debe querer decir que consideras que lo determinante para zanjar la cuestión es su conducta verbal de reportes, pero entonces descartas *a priori* la posibilidad del trastorno que describí como hipocondría de la visión en color.

No hay perspectivas de *encontrar* las respuestas a estos juegos de ingenio en nuestros usos cotidianos o en las intuiciones que éstos generan, pero desde luego el filósofo puede *crear* un edificio teórico para defender un conjunto particular de propuestas entrelazadas. El problema consiste en que, aunque cierta familia de estímulos y condiciones corporales produce normalmente una cierta familia de efectos, puede desconectarse cualquier efecto particular, y nuestras intuiciones no nos dicen cuáles son los efectos "esenciales" a la identidad del *quale* o a la constancia de los *qualia* (*cfr.* Dennett 1978a, cap. 11). Me parece bastante obvio que ninguno de los problemas reales de interpretación que confrontamos en estos casos curiosos se soluciona mediante un análisis de la forma en que debe aplicarse el concepto de *qualia*, a menos que queramos proponer un sentido nuevo y técnico para el cual el término tradicional se adoptaría. Sin embargo, eso sería, como mínimo, un error táctico: las intuiciones que rodean y *pretenden* anclar la comprensión actual del

término demuestran hallarse en completo desorden cuando se las confronta con estos casos.

El muestreo informal que he presentado hace patente que algunos filósofos poseen opiniones firmes acerca de cada caso y de la manera en que tendría que ser descrito cada uno en términos de *qualia*, pero encuentran que están en franco (y en última instancia, cómico) desacuerdo con otros filósofos sobre la manera en que tienen que hacerse estas descripciones "obvias". Otros filósofos descubren que no saben realmente qué decir, no porque no se hayan presentado suficientes hechos en las descripciones de los casos, sino porque comienza a volverse patente para ellos que no han sabido a ciencia cierta de qué han hablado a lo largo de los años.

6. *Llenar el vacío*

Si los *qualia* constituyen una pésima idea, ¿por qué han parecido ser una idea tan buena? ¿Por qué parece que existen esas propiedades intrínsecas, inefables, privadas y "cualitativas" en nuestra experiencia? Una revisión de las presuntas propiedades de segundo orden de las propiedades de nuestras experiencias conscientes nos permitirá diagnosticar su atractivo y encontrar sustitutos adecuados (para un ejercicio similar, véase Kitcher (1979)).

Consideremos primero lo "intrínseco". Está lejos de resultar claro lo que sería una propiedad intrínseca. Aunque el término ha gozado de cierta moda en filosofía, y a menudo parece asegurar un contraste importante, nunca ha habido una definición aceptada de esa propiedad de segundo orden de intrinsecalidad. Si hasta un proponente de teorías tan brillante como David Lewis ha intentado definir coherentemente la distinción entre lo extrínseco y lo intrínseco, y ha fracasado en ello, según él mismo lo ha admitido, podemos comenzar a preguntarnos si el concepto merece después de todo que le prestemos mayor atención. De hecho, Lewis (1983) empieza su estudio sobre las versiones de esa distinción listando como una de las opciones: "Podríamos quinear el conjunto, tachando a todo el grupo de ininteligible y prescindible", pero descarta esta sugerencia de inmediato: "Eso sería absurdo" (p. 197). Al final, no obstante,

su esfuerzo por salvar las explicaciones de Chisholm (1976) y Kim (1982) llega a un callejón sin salida, y conjetura que "si aun queremos irrumpir, lo mejor será intentarlo por otra ventana" (p. 200).

Aun cuando nos sintamos tan renuentes como Lewis a abandonar la distinción, ¿no debería despertar nuestras sospechas el siguiente hecho curioso? Si a un neófito se lo exhortara a explicar la idea de lo que es una propiedad intrínseca, muchas personas recurrirían al siguiente tipo de ejemplo: consideremos la pelota de Tomás; tiene muchas propiedades, como la de estar hecha de hule de la India, de pertenecer a Tomás, de haber pasado la última semana en el ropero y de ser roja. Todas, excepto la última, son propiedades claramente *relacionales* o *extrínsecas* de la pelota. Su color rojo, no obstante, constituye una propiedad intrínseca. Salvo que no es así. Desde Boyle y Locke tenemos un mayor conocimiento al respecto. El rojo —el rojo público— es una propiedad esencialmente relacional, como lo demuestran muchos experimentos mentales acerca de las "cualidades secundarias". (Uno de los primeros fue el cubo de agua templada de Berkeley (1713), y uno de los mejores es el de la feniltiourea de Bennett (1965)). El paso seductor, al saber que el rojo público (como el amargor público, etc.) es después de todo una propiedad relacional, consiste en aferrarse a la cualidad de intrinsecalidad ("*algo* tiene que ser intrínseco") e introducirla en la cabeza del sujeto. Suele creerse, de hecho, que si asumimos una postura relacional lockeana sobre el amargor, el rojo y demás cualidades objetivas, *tenemos* que completar nuestra explicación de las relaciones en cuestión recurriendo a propiedades intrínsecas, no relacionales. Si lo que es ser objetivamente amargo es producir un cierto efecto en los miembros de la clase de los observadores normales, tenemos que ser capaces de especificar ese efecto y distinguirlo del efecto producido por la acidez objetiva y demás.

¿Qué otra cosa serviría para distinguir este efecto sino una propiedad intrínseca? ¿Por qué no otra propiedad relacional o extrínseca? El tratamiento relacional del valor monetario no exige, para ser completado, que se supongan elementos de valor intrínseco (independiente en valor de las disposiciones de los valuadores a reaccionar conductualmente). La afirmación

de que ciertas propiedades perceptuales son diferentes resulta, en ausencia de un argumento que la sustente, una petición de principio. No servirá de nada decir que es obvio que son intrínsecas. Pudo parecerle obvio a algunas personas, pero las consideraciones planteadas por el dilema de Chase muestran que está lejos de ser obvio que cualquier propiedad intrínseca (independientemente de lo que esto sea) puede desempeñar el papel de ancla en el tratamiento relacional lockeano de las propiedades perceptuales públicas.

¿Por qué no abandonar por completo la intrinsecalidad como propiedad de segundo orden, por lo menos hasta que se resuelva el desorden de la opinión filosófica sobre lo que podría ser la intrinsecalidad? Hasta ese momento, insistir en que los *qualia* son las propiedades intrínsecas de la experiencia constituye, a lo sumo, un gesto vacuo; nadie podría afirmar que ofrece un prerrequisito comprendido, claro y coherente de una teoría.[12]

¿Qué decir entonces de la inefabilidad? ¿Por qué parece que nuestras experiencias conscientes poseen propiedades inefables? Porque, en efecto, tienen propiedades inefables *en la práctica*. Supongamos, en el *generador de intuiciones 13: el chillido del pigargo*, que jamás he oído el chillido del pigargo, ni siquiera grabado, pero tengo la idea general, después de leer mis libros sobre aves, de a qué debo prestar atención: "una serie de silbidos cortos y agudos, como chirridos: *chip, chip*, o *chuc, chuc*, etc.; suena enojado" (Peterson 1947) (o palabras semejantes o más precisas). La descripción verbal me da un límite parcial en el espacio lógico de los posibles chillidos de aves. Sobre esta base puedo descartar muchos chillidos que he escuchado o podría escuchar, pero sigue habiendo una amplia gama de posibilidades que puedo discriminar y dentro de la cual la realidad se oculta de mí como una aguja en un pajar.

Entonces un día, armado con mi descripción verbal y con mis binoculares, identifico visualmente un pigargo y luego oigo su

[12] Una negación heroica (y, en mi opinión, desconcertante) a abandonar la intrinsecalidad es la contemplación de Sellars (1981) a través de los años de su famoso cubo de hielo rosa, la cual lo lleva a postular una revolución en microfísica, restableciendo los "procesos sensoriales absolutos" objetivos en contra de Boyle y Locke, y de casi todo el mundo después de ellos (véase también mi comentario en Dennett (1981)).

chillido. "De modo que así es como suena", me digo, ostentando –al parecer– un complejo mental particular de *qualia* intrínsecos e inefables. Nombro al complejo "*S*" (con perdón de Wittgenstein), lo repaso en la memoria a corto plazo, lo cotejo con las descripciones de los libros de aves y veo que, si bien las descripciones verbales son verdaderas, exactas y hasta poéticamente evocadoras –decido que yo no lo podría describir mejor ni con un millar de palabras–, aún no logran *captar* el complejo de *qualia* que he llamado *S*. De hecho, ésta es la razón de que necesite el neologismo, "*S*", para referir directamente a la propiedad inefable a la que no puedo denotar con la descripción. Mi experiencia perceptual ha identificado para mí la localización del chillido del pigargo en el espacio lógico de posibilidades de una manera que la descripción verbal no lo hizo.

Pero, pese a lo tentadora que es esta perspectiva de la cuestión, resulta exagerada. En primer lugar, es obvio que, a partir de una única experiencia de este tipo, yo no sé –no puedo saber– cómo generalizar a otros chillidos del pigargo. Un chillido que sólo se diferenciara por ser media octava más alto, ¿también sería un chillido de pigargo? Ésta es una cuestión empírica ornitológica para la que mi experiencia ofrece una escasa evidencia. Pero además –y esto es un asunto psicológico, no ornitológico–, no sé ni puedo saber, a partir de esa única experiencia, qué variaciones y constancias físicas, en estímulos, producirían en mí una experiencia indistinguible. Tampoco puedo saber si reaccionaría igual (si tendría la misma experiencia) en caso de que me presentaran algo que fuera, desde el punto de vista de todas las medidas físicas, una segunda estimulación idéntica a la primera. No puedo saber el efecto modulador, si lo hay, de las variaciones que se producen en mi cuerpo (o en mi psique).

Esta inescrutabilidad de la proyección es sin duda una de las fuentes de la verosimilitud del escepticismo de Wittgenstein con respecto a la posibilidad de un lenguaje privado:

> Wittgenstein enfatiza que las definiciones ostensivas en principio siempre pueden ser mal comprendidas, inclusivo la definición ostensiva de una palabra de color como "sepia". Cómo entiende

> alguien la palabra es algo que queda exhibido en la forma en que prosigue, "el uso que hace de la palabra definida". Se puede ir por el buen camino dada una explicación mínima, en tanto que, por otra parte, se puede ir por otro camino independientemente de cuántas aclaraciones se añadan, puesto que éstas también pueden ser incomprendidas. (Kripke 1982, p. 83; véanse también las pp. 40–46.)

Pero lo que es inescrutable a simple vista, y un poco ambiguo después de pruebas limitadas, puede llegarse a entender justificablemente como el resultado de un detector de propiedades altamente específico, confiable y proyectable, una vez que se lo ha sometido a pruebas de campo bajo una variedad de circunstancias adecuadamente amplia.

En otras palabras, cuando oigo por primera vez el chillido del pigargo, quizá identifico un detector de propiedades en mí mismo, pero no tengo (todavía) una idea de cuál es la propiedad que detecta mi recién descubierto detector de propiedades. Podría parecer, entonces, que no sé absolutamente nada nuevo, que mi novedosa experiencia no ha mejorado mi predicamento epistémico en lo más mínimo. Pero, desde luego, no es así. Quizá no soy capaz de describir la propiedad o de identificarla en relación con algún rasgo público al que pueda recurrir fácilmente (todavía), pero estoy familiarizado con ella en un grado modesto: puedo referir a la propiedad que detecté: es la propiedad que detecté en *ese* suceso. Mi experiencia del chillido del pigargo me ha brindado una nueva forma de pensar acerca de los chillidos de los pigargos (una manera inevitablemente pomposa de expresar algo muy sencillo), la cual es prácticamente inefable tanto porque posee (todavía para mí) un perfil no probado en respuesta a circunstancias perceptuales, como porque es —como lo atestigua la pobreza de la descripción del libro sobre aves— una forma sumamente informativa de pensar: un resultado de una parte informacionalmente muy sensible de mi sistema nervioso.

En este caso hablo de información en el sentido formal de la teoría de la información. Consideremos (*generador de intuición 14: la caja de gelatina Jello*) el viejo truco de los espías, cuya exposición más conocida es el caso de Julius y Ethel Rosenberg,

en el cual se mejora un sistema de contraseñas partiendo algo en dos (una caja de gelatina Jello, en el caso de los Rosenberg) y dando una mitad a cada una de las partes, que tienen que tener cuidado en identificarse mutuamente. ¿Por qué funciona? Porque rasgar el papel en dos produce un borde de tal complejidad informativa que sería casi imposible reproducirlo mediante una construcción deliberada. (Cortar la caja de gelatina por un borde recto con una navaja frustraría el propósito por completo.) El borde serrado particular de un pedazo se convierte en un mecanismo *prácticamente* único de reconocimiento de patrones para su contrapartida; se trata de un mecanismo para detectar la propiedad de forma *M*, donde *M* es únicamente instanciada por su contrapartida. La esencia del truco es que no podemos sustituir nuestro predicado simulado "*M*" por una descripción más extensa y compleja, pero precisa y exhaustiva, de la propiedad, porque, si pudiéramos hacerlo, podríamos utilizar la descripción como receta o algoritmo factible para producir otra instancia de *M* u otro detector de *M*. La única manera *fácilmente accesible* de decir lo que es la propiedad *M* consiste en señalar nuestro detector de *M* y decir que *M* es la propiedad de forma detectada por esta cosa que está aquí.

Y eso es precisamente lo que hacemos cuando parecemos ostentar, con el dedo mental de la intención interna, un *quale* o un complejo de *qualia* en nuestra experiencia. Nos referimos a una propiedad —una propiedad pública de fronteras indefinidas— haciendo referencia a nuestra capacidad personal e idiosincrásica para responder a ella. Esa idiosincrasia es el alcance de nuestra vida privada. Si pregunto si tu azul es mi azul, si tu do mayor es mi do mayor, puedo estar preguntando, coherentemente, si nuestros perfiles discriminatorios en una amplia variación de condiciones serán más o menos los mismos. Y pueden no serlo; las personas experimentan el mundo de un modo muy diferente. Pero esto puede descubrirse empíricamente mediante todos los procedimientos habituales de pruebas objetivas.[13]

[13] Stich (1983) discute las implicaciones, para la teoría psicológica, de los problemas de inconmensurabilidad que pueden surgir de tales diferencias en los perfiles de discriminación (véanse esp. los caps. 4 y 5).

Peter Bieri me ha señalado que hay una forma natural de explotar el sentido de Dretske (1981) de información en una reformulación de mis tres primeras propiedades de segundo orden de los *qualia*: intrinsecalidad, inefabilidad y privacía. (El intento de Dretske por utilizar la teoría de la información de este modo presenta problemas –véase mi discusión en Dennett 1987b, cap. 8–, pero no son importantes para esta cuestión.) Podríamos hablar de lo que Bieri llamaría "propiedades de información fenoménica" de los sucesos psicológicos. Consideremos la información –lo que Dretske llamaría el significado natural– que podría llevar un tipo de suceso perceptual interno. El que lleve esa información es una cuestión objetiva (y por lo tanto, en un sentido laxo, intrínseco) porque es independiente de qué información (si es que la hay) el sujeto *cree* que lleva el tipo de suceso. Cuál es exactamente la información que lleva resulta (prácticamente) inefable, por las razones que se acaban de dar. Y es privada en el sentido que se acaba de exponer: implica pertenencia y es potencialmente idiosincrásica.

Consideremos la manera en que se aplicarían las "propiedades de información fenoménica" propuestas por Bieri (llamémoslas *pifs*) al caso de Chase y Sanborn. Tanto Chase como Sanborn deberían preguntarse si sus *pifs* han cambiado. El discurso de Chase muestra que tiene la impresión de que sus *pifs* permanecen inalteradas (en circunstancias normales; todas las apuestas se desploman si acaba de comer rábanos picantes). Cree que las mismas cosas objetivas del mundo –en particular, los líquidos con un alto contenido de cafeína que son químicamente idénticos– dan origen a sus tipos particulares de experiencias gustativas tanto ahora como hace seis años.

Sanborn tiene la impresión de que sus *pifs* son diferentes. Piensa que sus detectores de propiedades objetivas están trastornados. Ya no tiene confianza en que los resultados que arrojan hoy le informen lo mismo que hace seis años. ¿Y qué le informaban exactamente en ese entonces? Si Sanborn fuera una persona ordinaria, no esperaríamos que tuviera una respuesta explícita, porque la mayoría de nosotros tratamos a nuestros detectores gustativos como simples detectores *M*, detectando lo que sea que detectan. (Hay buenas razones para esto, las cuales han sido analizadas por Akins, tesis inédita, 1987.) Pero los

catadores de café profesionales son probablemente distintos. Probablemente tienen una idea muy clara del tipo de maquinaria de transducción de análisis químicos que tienen en la boca y en el sistema nervioso.

Hasta aquí, todo bien. Podríamos reinterpretar los discursos de Chase y de Sanborn como hipótesis sobre las constancias o los cambios en los resultados de sus aparatos procesadores de información perceptual, y precisamente el tipo de pruebas empíricas que imaginamos antes tendería a confirmar o desconfirmar sus opiniones así interpretadas. Pero, ¿qué justificaría llamar "fenoménica" a esa propiedad portadora de información?

Tal *pif* posee, como lo revela el testimonio de Chase y Sanborn, el poder de provocar en ellos actos de (aparente) reidentificación o reconocimiento. Este poder es, desde luego, una propiedad disposicional lockeana equiparable al poder de las cosas amargas de provocar cierta reacción en las personas. Es este poder solo, independientemente de cómo se realice en el cerebro, lo que da a Chase y a Sanborn "acceso" a los resultados de sus respectivos detectores de propiedades.

Podemos "señalar interiormente" uno de los resultados de nuestros detectores de propiedades propios e idiosincrásicos, pero al hacerlo, ¿qué estamos señalando? ¿*En qué consiste* ese resultado en sí mismo? ¿O cuáles son sus propiedades conscientemente aprehensibles, si no justamente nuestros proscritos amigos, los *qualia*? Debemos tener cuidado en este punto, pues si invocamos un proceso perceptual interno en el que observamos el resultado con algún ojo interior y, de este modo, discernimos sus propiedades, estaremos cayendo otra vez en la concepción problemática de que los *qualia* no son más que propiedades comunes de nuestros estados internos.

Pero nada nos exige hacer esa invocación. No tenemos que saber cómo identificar o reidentificar o acceder a tales tipos de respuestas internas a fin de ser capaces de identificarlos. Ésta es una observación presentada con fuerza por los primeros funcionalistas y materialistas, y nunca ha sido rebatida (Farrell 1950; Smart 1959). Las propiedades de la "cosa experimentada" no deben confundirse con las propiedades del suceso que realiza

el experimentarla. Para expresarlo en términos vívidos, la diferencia física entre que alguien imagine una vaca púrpura y que imagine una vaca verde *podría* no ser más que la presencia o la ausencia de un cero o un uno particulares en alguno de los "registros" del cerebro. Tal presencia física bruta es todo lo que se necesitaría para anclar las clases de diferencias disposicionales entre imaginar una vaca púrpura e imaginar una vaca verde, las cuales luego fluirían causalmente de ese hecho "intrínseco". (Dudo que esto sea lo que han tenido en mente los partidarios de los *qualia* cuando insisten en que los *qualia* son propiedades intrínsecas.)

Además, es nuestra incapacidad misma para expandir, o modificar, estas disposiciones brutas a fin de identificar o reconocer dichos estados lo que crea la ilusión doctrinal de "homogeneidad", "atomicidad frente al análisis" o "uniformidad" que caracteriza a los *qualia* de la tradición filosófica.

Mi hipótesis consiste en que esta supuesta uniformidad no es más que una especie de invariabilidad funcional: un pariente cercano de lo que Pylyshyn (1980, 1984) llama *impenetrabilidad cognoscitiva*. Además, esta invariabilidad funcional o impenetrabilidad no es absoluta, sino que resulta ser ella misma plástica a través del tiempo. Tal como, en la parte eferente del sistema nervioso, se ha descubierto que las *acciones básicas* —en el sentido que les dan Danto (1963, 1965) y otros (véase Goldman 1970)— son variables, y que bajo entrenamiento están sujetas a descomposición (se puede aprender, con la ayuda de la "biorretroalimentación", a provocar voluntaria y "directamente" la activación de una neurona motora particular), así que lo que cuenta para un individuo como las propiedades simples o atómicas de las cosas experimentadas está sujeto a variación mediante entrenamiento.[14]

Consideremos los resultados de "educar" el paladar de un catador de vino o de "entrenar el oído" de los músicos. Lo que había sido "atómico" o "inanalizable" se vuelve claramente

[14] Véase Churchland (1979, esp. el cap. 2) para observaciones en apoyo de la variabilidad de las propiedades perceptuales, y para argumentos nuevos contra el uso de "propiedades intrínsecas" como determinantes del significado de los predicados perceptuales. Véase también Churchland (1985) para argumentos y observaciones adicionales en apoyo de la postura aquí esbozada.

compuesto y descriptible; pares que habían sido indistinguibles pueden ahora distinguirse, y cuando esto sucede decimos que *la experiencia cambia*. Un ejemplo rápido y revelador de esto se ilustra en el *generador de intuición 15: la cuerda de la guitarra*. Puntea la cuerda mi baja o grave sola, y escucha el sonido atentamente. ¿Tiene partes descriptibles o se trata de uno solo e inefablemente guitarresco? Muchos optarán por decir lo segundo. Vuelve ahora a puntear la cuerda sola y lleva con cuidado un dedo al octavo traste, presionándolo ligeramente para crear un sonido "armónico" agudo. De pronto se escucha un sonido *nuevo*: de alguna manera "más puro" y, desde luego, una octava más alto. Algunas personas insistirán en que se trata de un sonido completamente nuevo, mientras que otras describirán la experiencia diciendo que "la base se separó de la nota", dejando sólo la parte alta. Pero luego, con un tercer punteado, se puede oír, con sorprendente nitidez, el sobretono armónico que se aisló la segunda vez que se punteó la cuerda. La homogeneidad y la inefabilidad de la primera experiencia han desaparecido, y han sido reemplazadas por una dualidad tan "directamente aprehensible" y claramente descriptible como la de cualquier cuerda.

La diferencia en la experiencia es asombrosa, pero la complejidad aprehendida en el tercer punteado estuvo *ahí* todo el tiempo (al responder a ella o al discriminarla). Después de todo, fue por el complejo patrón de sobretonos que se pudo reconocer el sonido como el de una guitarra, y no como de un laúd o un clavicordio. En otras palabras, aunque la experiencia subjetiva ha cambiado drásticamente, la *pif* no ha cambiado; uno sigue respondiendo, como antes, a una propiedad compleja tan altamente informativa que prácticamente desafía toda descripción verbal.

No hay nada que detenga el refinamiento ulterior de la capacidad de uno para describir esta complejidad que, hasta ahora, es inefable. En cualquier momento dado, por supuesto, se encuentra el propio horizonte actual de distinguibilidad, y ese horizonte es lo que establece, si acaso algo lo hace, lo que tendríamos que llamar las propiedades primarias o atómicas de lo que uno experimenta conscientemente (Farrell 1950). Pero sería un error transformar el hecho de que inevitablemente

hay un límite para nuestra capacidad de describir las cosas que experimentamos en la suposición de que nuestra experiencia contiene propiedades absolutamente indescriptibles.

Así, cuando le echamos una última mirada a nuestra caracterización original de los *qualia* como propiedades inefables, intrínsecas, privadas y directamente aprehensibles de la experiencia, descubrimos que no hay nada que llene esos requisitos. En su lugar encontramos propiedades públicas relativa o prácticamente inefables a las que podemos referirnos de manera indirecta por referencia a nuestros detectores de propiedades privadas —sólo en el sentido de que son idiosincrásicas. Y en la medida en que pretendamos aferrarnos a nuestra autoridad subjetiva sobre la presencia, dentro de nosotros, de estados de cierto tipo o con ciertas propiedades, podemos tener alguna autoridad —no infalibilidad o incorregibilidad, sino algo mejor que mera adivinanza—, pero sólo si nos restringimos a propiedades relacionales, extrínsecas, como el poder de ciertos estados internos nuestros de provocar actos de aparente reidentificación. Así, contrario a lo que a primera vista parece obvio, simplemente no hay *qualia*.

[*Traducción de Ana Isabel Stellino*]

BIBLIOGRAFÍA

Akins, K., 1987, "Information and Organisms: or Why Nature Doesn't Build Epistemic Engines" (tesis de doctorado inédita), University of Michigan, Ann Arbor.

Armstrong, D. y N. Malcolm (comps.), 1984, *Consciousness and Causality*, Blackwell Scientific Publications, Oxford.

Bennett, J., 1965, "Substance, Reality and Primary Qualities", *American Philosophical Quarterly*, vol. 2, pp. 1-17.

Berkeley, G., 1713, *Three Dialogues between Hylas and Philonous*, Londres. [Versión en castellano: *Tres diálogos entre Hilas y Filonus en oposición a escépticos y ateos*, trad. A.P. Masegosa, Aguilar, Buenos Aires, 1978.]

Block, N., 1980, "Are Absent Qualia Impossible?", *Philosophical Review*, vol. 89, pp. 257-274.

Block, N., 1978, "Troubles with Functionalism", en Savage 1978, pp. 261–326.
——— y J. Fodor, 1972, "What Psychological States Are Not", *Philosophical Review*, vol. 81, pp. 159–181. [Incluido en esta antología, pp. 155–181.]
Butler, R.J. (comp.), 1965, *Analytical Philosophy*, segunda serie, Blackwell, Oxford.
Chisholm, R., 1976, *Person and Object*, Open Court Press, La Salle.
Churchland, P.M., 1985, "Reduction, Qualia and the Direct Inspection of Brain States", *Journal of Philosophy*, vol. 2, no. 1, pp. 8–28.
———, 1979, *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge.
Conee, E., 1985, "The Possibility of Absent Qualia", *Philosophical Review*, vol. 94, no. 3, pp. 345–366.
Damasio, A., T. Yamada, H. Damasio, J. Corbett y J.McKee, 1980, "Central Achromatopsia: Behavioral, Anatomic, and Physiological Aspects", *Neurology*, vol. 30, pp. 1064–1071.
Danto, A., 1965, "Basic Actions", *American Philosophical Quarterly*, vol. 60, pp. 141–148.
———, 1963, "What We Can Do", *Journal of Philosophy*, vol. 60, pp. 435–445.
Davis, L., 1982, "Functionalism and Absent Qualia", *Philosophical Studies*, vol. 41, no. 2, pp. 231–251.
———, 1987a, *The Intentional Stance*, MIT Press, Cambridge, Mass. [Versión en castellano: *La actitud intencional*, trad. D. Zadunaisky, Gedisa, Barcelona, 1991.]
Dennett, D.C., 1987b, *The Philosophical Lexicon* (8a. ed.). Copia disponible en la Asociación Filosófica Estadounidense, University of Delaware, Newark.
———, 1982, "How to Study Human Consciousness Empirically: Or Nothing Comes to Mind", *Synthese*, vol. 53, pp. 159–180.
———, 1981, "Wondering Where the Yellow Went", *Monist*, vol. 64, pp. 102–108.
———, 1979, "On the Absence of Phenomenology", en Gustafson y Tapscott 1979, pp. 93–114.
———, 1978a, *Brainstorms*, MIT Press, Cambridge, Mass.
———, 1978b, "Two Approaches to Mental Images", en Dennett 1978a, pp. 174–189.
———, 1976, "Are Dreams Experiences?", *Philosophical Review*, vol. 73, pp. 151–171.
———, 1969, *Content and Consciousness*, Routledge and Kegan Paul, Londres. [Versión en castellano: *Contenido y conciencia*, trad. J.M. Lebrón, Gedisa, Barcelona, 1996.]

Dretske, F., 1981, *Knowledge and the Flow of Information*, MIT Press, Cambridge, Mass. [Versión en castellano: *Conocimiento e información*, trad. M. Vicedo, M. Guilla y F. Pizarro, Salvat, Barcelona, 1989.]

Elster, J., 1985, *Making Sense of Marx*, Cambridge University Press, Cambridge.

Farrell, B., 1950, "Experience", *Mind*, vol. 59, pp. 170–198.

Gert, B., 1965, "Imagination and Verifiability", *Philosophical Studies*, vol. 16, pp. 44–47.

Geschwind, N. y M. Fusillo, 1966, "Color-Naming Defects in Association with Alexia", *Archives of Neurology*, vol. 15, pp. 137–146.

Goldman, A., 1970, *A Theory of Human Action*, Prentice-Hall, Englewood Cliffs.

Gregory, R., 1977, *Eye and Brain*, 3a. ed., Weidenfeld and Nicolson, Londres.

Gustafson, D.F. y B.L. Tapscott (comps.), 1979, *Body, Mind, and Method*, Reidel, Dordrecht.

Hofstadter, D. y D.C. Dennett, 1981, *The Mind's I: Fantasies and Reflections on Self and Soul*, Basic Books, Nueva York.

Kim, J., 1982, "Psychophysical Supervenience", *Philosophical Studies*, vol. 41, pp. 51–70.

Kitcher, P., 1979, "Phenomenal Qualities", *American Philosophical Quarterly*, vol. 16, pp. 123–129.

Kripke, S., 1982, *Wittgenstein on Rules and Private Language*, Harvard University Press, Cambridge, Mass.

Lehrer, A., 1983, *Wine and Conversation*, University of Indiana Press, Bloomington.

Lewis, D., 1983, "Extrinsic Properties", *Philosophical Studies*, vol. 44, pp. 197–200.

Locke, J., 1690, *An Essay Concerning Human Understanding*, ed. A.C. Fraser, Dover, Nueva York (edición de 1959).

Lycan, W., 1973, "Inverted Spectrum", *Ratio*, vol. 15, pp. 315–319.

Malcolm, N., 1956, "Dreaming and Skepticism", *Philosophical Review*, vol. 65, pp. 14–37.

———, 1959, *Dreaming*, Routledge and Kegan Paul, Londres.

Meadows, J.C., 1974, "Disturbed Perception of Colours Associated with Localized Cerebral Lesions", *Brain*, vol. 97, pp. 615–632.

Millikan, R., 1984, *Language, Thought and Other Biological Categories*, MIT Press, Cambridge, Mass.

Nagel, T., 1974, "What Is It Like to Be a Bat?", *Philosophical Review*, 83, pp. 435–451. [Incluido en esta antología, pp. 45–63.]

Nagel, T., 1986, *The View from Nowhere*, Oxford University Press, Oxford. [Versión en castellano: Una visión de ningún lugar, trad. Fondo de Cultura Económica, México, 1996.]

Peirce, C., 1931–1958, *Collected Works*, vol. 5, C. Hartshorne y P. Weiss, Harvard University Press, Cambridge, Mass.

Peterson, R.T., 1947, *A Field Guide to the Birds*, Houghton Mifflin, Boston.

Putnam, H., 1965, "Brains and Behavior", en Butler 1965, pp. 1–19.

Pylyshyn, Z., 1980, "Computation and Cognition: Issues in the Foundations of Cognitive Science", *Behavioral and Brain Sciences*, vol. 3, pp. 111–132. [Versión en castellano: *Computación y conocimiento: Hacia una fundamentación de la ciencia cognitiva*, trad. R. Fernández González, Debate, Madrid, 1988.]

———, 1984, *Computation and Cognition: Toward a Foundation for Cognitive Science*, MIT Press, Cambridge, Mass.

Savage, C.W. (comp.), 1978, *Perception and Cognition: Issues in the Foundations of Psychology*, University of Minnesota Press, Minneápolis.

Sellars, W., 1981, "Foundations for a Metaphysics of Pure Process" (Conferencias Carus), *Monist*, vol. 64, pp. 3–90.

Shoemaker, S., 1984, *Identity, Cause, and Mind*, Cambridge University Press, Cambridge, pp. 351–357.

———, 1982, "The Inverted Spectrum", *Journal of Philosophy*, vol. 79, pp. 357–381.

———, 1981, "Absent Qualia Are Impossible —A Reply to Block", *Philosophical Review*, vol. 90, pp. 581–599.

———, 1975, "Functionalism and Qualia", *Philosophical Studies*, vol. 27, pp. 291–315. [Incluido en esta antología, pp. 183–212.]

———, 1969, "Time Without Change", *Journal of Philosophy*, vol. 66, pp. 363–381.

Smart, J.C., 1959, "Sensations and Brain Processes", *Philosophical Review*, vol. 68, pp. 141–156.

Stich, S., 1983, *From Folk Psychology to Cognitive Science: The Case against Belief*, MIT Press, Cambridge, Mass.

Taylor, D.M., 1966, "The Incommunicability of Content", *Mind*, vol. 75, pp. 527–541.

Von der Heydt, R., E. Peterhans, y G. Baumgarther, 1984, "Illusory Contours and Cortical Neuron Response", *Science*, vol. 224, pp. 1260–1262.

Walzer, M., 1985, "What's Left of Marx", *New York Review of Books*, 21 de noviembre, pp. 43–46.

White, S., 1985, "Professor Shoemaker and So-Called 'Qualia' of Experience", *Philosophical Studies*, vol. 47, pp. 369–383.

Wittgenstein, L., 1958, *Philosophical Investigations*, Blackwell Scientific Publications, Oxford. [Versión en castellano: *Investigaciones filosóficas*, trad. A. García Suárez y U. Moulines, Crítica/Instituto de Investigaciones Filosofícas, Barcelona, 1988, p. 237, §279.]

LA CUALIDAD INTRÍNSECA DE LA EXPERIENCIA*

GILBERT HARMAN

1. *El problema*

Muchos filósofos, psicólogos e investigadores en inteligencia artificial aceptan una visión ampliamente funcionalista de la relación entre mente y cuerpo, considerando, por ejemplo, que la mente es en el cuerpo algo parecido a una computadora en un robot, quizá con un procesamiento masivamente paralelo (como en Rumelhart y McClelland 1986). Pero a esta idea de la mente no le han faltado impugnaciones. Algunos filósofos y otros objetan enérgicamente que el funcionalismo debe inevitablemente fracasar al dar cuenta de la parte más importante de la vida mental, a saber, la sensación subjetiva de la experiencia consciente.

El modelo computacional de la mente representa una versión del funcionalismo, aunque no es la única versión. En su forma más general, el funcionalismo define los estados y los procesos mentales en términos de las relaciones causales o funcionales que guardan entre sí y con las entradas perceptuales provenientes del mundo exterior y las salidas conductuales expresadas en acciones. De acuerdo con el funcionalismo, lo importante son las relaciones funcionales, no las cualidades intrínsecas de

* La preparación de este escrito fue apoyada, en parte, por las becas de investigación otorgadas a Princeton University por la Fundación James S. McDonnell y la Fundación Nacional para las Ciencias.

las cosas en las cuales estas relaciones se instancian. Del mismo modo en que los mismos programas computacionales pueden correr en computadoras diferentes construidas con materiales distintos, el funcionalismo permite que los mismos estados y sucesos mentales se den en seres con constituciones físicas muy diversas, ya que exactamente las mismas relaciones funcionales pueden instanciarse en seres con configuraciones físicas muy distintas. Según el funcionalismo, creencias, deseos, pensamientos y sensaciones no se limitan a seres que materialmente son como nosotros. Esos estados y sucesos psicológicos también pueden ocurrir, por ejemplo, en seres conformados con base en el silicio, siempre que se den las relaciones funcionales correctas.

El funcionalismo puede permitir la posibilidad de que algo acerca del silicio imposibilite que se den las relaciones pertinentes en los seres de silicio, quizá debido a que los sucesos pertinentes no pueden ocurrir con suficiente rapidez en el silicio. Es incluso concebible que las relaciones funcionales pertinentes sólo pudieran darse en el tipo de material que compone el cerebro humano (Thagard 1986; Dennett 1987, cap. 9). El funcionalismo implica que en tal caso el material es importante sólo porque se lo necesita para las relaciones funcionales pertinentes y no debido a alguna otra vinculación más misteriosa o mágica entre ese tipo de materia y cierta clase de conciencia.

Surgen varias cuestiones dentro del enfoque funcionalista general. Para comenzar, hay una polémica acerca de la manera de identificar las entradas de un sistema funcional. ¿Deben identificarse esas entradas con sucesos del entorno externo (Harman 1988) o deben, más bien, ser identificadas con sucesos que son más internos, como la estimulación de los órganos sensoriales de un organismo (Block 1986)? También existe la posibilidad de desacuerdo en cuanto a qué tan deterministas deben ser las relaciones funcionales pertinentes. ¿Deben ser completamente deterministas, o pueden ser meramente probabilísticas? ¿O podrían incluso simplemente ser no deterministas, sin siquiera asociarse a probabilidades definidas (Harman 1973, pp. 51–53)?

No me ocuparé aquí de estos asuntos. En cambio me concentraré en la cuestión diferente y más básica que ya he mencio-

nado, a saber, si esta clase de funcionalismo, no importa cuán elaborado sea, puede dar cuenta de la sensación subjetiva de la experiencia, del "cómo es" (Nagel 1974) tener tal o cual experiencia. Aun más, no consideraré la impugnación general de "cómo el funcionalismo da cuenta de *X*" para tal o cual *X*. Tampoco voy a considerar los argumentos negativos contra análisis funcionalistas particulares. Más bien consideraré tres argumentos relacionados que pretenden demostrar que el funcionalismo no puede dar cuenta de este aspecto de la experiencia. Sostendré que estos tres argumentos son falaces. Poco de lo que diré será original, y en gran medida sólo desarrollaré ideas presentadas hace muchos años (Quine 1960, p. 235, Anscombe 1965, Armstrong 1961, 1962 y en especial 1968, Pitcher 1971), ideas que, según creo, no han sido apreciadas de manera adecuada. Los tres argumentos son los siguientes:

> Primero, cuando uno presta atención a un dolor en la pierna o a su experiencia de lo rojo de una manzana, es consciente de una cualidad intrínseca de su experiencia, donde una cualidad intrínseca es una cualidad que algo tiene en sí mismo, aparte de sus relaciones con otras cosas. Esta cualidad de la experiencia no puede captarse en una definición funcional, ya que a tal definición le conciernen solamente las relaciones, relaciones entre los estados mentales y la entrada perceptual, relaciones de los estados mentales entre sí y relaciones entre los estados mentales y la salida conductual. Por ejemplo, "Una característica esencial del análisis [funcionalista de Armstrong] es que no nos dice nada sobre la naturaleza intrínseca de los estados mentales [...]. Nunca toma en serio la objeción natural de que debemos conocer la naturaleza intrínseca de nuestros propios estados mentales ya que los experimentamos directamente" (Nagel 1970).
>
> Segundo, una persona ciega de nacimiento podría saber todo acerca de los hechos físicos y funcionales de la percepción del color, sin saber cómo es ver algo rojo. Por lo tanto, cómo es ver algo rojo no puede explicarse en términos puramente funcionales (Nagel 1974, Jackson 1982, 1986).[1]
>
> Tercero, es concebible que dos personas tengan sistemas visuales que funcionan de manera similar, a pesar del hecho de que las

[1] Incluidos en esta antología: pp. 45-63, pp. 95-110, y pp. 111-118, respectivamente.

cosas que le parecen rojas a una persona le parecen verdes a la otra, las cosas que le parecen anaranjadas a la primera persona le parecen azules a la segunda, y así sucesivamente (Lycan 1973; Shoemaker 1982). Este tipo de inversión del espectro en la manera en que se ven las cosas es posible, pero no puede dársele una descripción puramente funcional, pues, por hipótesis, no existen diferencias funcionales entre las personas en cuestión. Puesto que la manera en que se ven las cosas para una persona es un aspecto de la vida mental de esa persona, esto significa que un aspecto importante de la vida mental de una persona no puede explicarse en términos puramente funcionales.

2. *Intencionalidad*

Con el fin de evaluar estos argumentos, comenzaré por comentar lo que en ocasiones se denomina la intencionalidad de la experiencia. Nuestra experiencia del mundo posee contenido —es decir, representa las cosas como siendo de cierta manera. En particular, la experiencia perceptual representa a un perceptor como situado en un entorno particular, por ejemplo, frente a un árbol de corteza café y hojas verdes que se agitan con una brisa ligera.

Una cosa que los filósofos quieren decir cuando hacen referencia a esto como el contenido intencional de la experiencia es que el contenido de la experiencia puede no reflejar lo que realmente está ahí. Aunque a mí me parece que estoy viendo un árbol, esto puede ser una hábil ilusión creada mediante espejos inclinados y telones de fondo pintados. O puede ser una alucinación producida por una droga en mi café.

Hay muchos otros ejemplos de intencionalidad. Ponce de León recorrió Florida en busca de la Fuente de la Juventud. Lo que buscaba era una fuente cuyas aguas le darían juventud eterna a quienquiera que las bebiera. En realidad, no hay tal cosa como la Fuente de la Juventud, pero eso no significa que Ponce de León no buscase nada. Estaba en busca de algo. En consecuencia, podemos decir que su búsqueda tenía un objeto intencional. Pero la cosa que estaba buscando, el objeto intencional de su búsqueda, no existía (y no existe).

Una pintura de un unicornio es una pintura de algo; posee cierto contenido. Pero el contenido no corresponde a nada real; la cosa representada por la pintura no existe. La pintura tiene un contenido intencional en el sentido pertinente de "intencional".

Resulta útil comparar imaginarse o representarse mentalmente un unicornio con la pintura de un unicornio. En ambos casos, el contenido no es real; el objeto representado, el objeto intencional de la representación, no existe. Es sólo un objeto intencional.

Esto no equivale a suponer que representarse mentalmente un unicornio involucra tener conciencia de una representación mental de un unicornio. Estoy comparando el representarse mentalmente una cosa con la pintura de una cosa, no con la percepción de una pintura. La conciencia de una pintura tiene como objeto intencional la pintura misma. La pintura tiene como objeto intencional un unicornio. Imaginar un unicornio es distinto de imaginar una pintura de un unicornio. El objeto intencional de ese imaginar es un unicornio, no una pintura de un unicornio.

Es muy importante distinguir entre las propiedades de un objeto representado y las propiedades de una representación de ese objeto. Claramente, estas propiedades pueden ser muy diferentes. El unicornio está representado como de cuatro patas y un solo cuerno. La pintura del unicornio no tiene cuatro patas y un solo cuerno. La pintura es plana y está cubierta de pintura. El unicornio no se representa como plano y cubierto de pintura. Asimismo, un unicornio imaginado se imagina como de patas y cuerno. El imaginarse el unicornio no tiene patas ni cuerno. El imaginarse el unicornio es una actividad mental. El unicornio no es imaginado ni como actividad ni como nada mental.

La conocida teoría de la percepción de los datos sensoriales surge a partir de la incapacidad de distinguir estas consideraciones elementales. De acuerdo con esa antigua teoría, la percepción de los objetos externos en el entorno es siempre indirecta y está siempre mediada por una conciencia más directa de un dato sensorial mental. Los defensores de la teoría de los datos sensoriales argumentan en favor de ella recurriendo al así denominado argumento de la ilusión. Este argumento parte de

la incontrovertida premisa de que la forma en que las cosas se presentan en la percepción no es siempre la forma en que son. Eloísa ve algo café y verde. Pero no hay nada café y verde ante ella; todo es una ilusión o una alucinación. A partir de esto, el argumento infiere falazmente que lo café y verde que ve Eloísa no es externo a ella y, por lo tanto, debe ser interno o mental. Puesto que la percepción verídica, no ilusoria ni alucinatoria puede ser cualitativamente indistinguible de la percepción ilusoria o alucinatoria, el argumento concluye que, en todos los casos de percepción, Eloísa es directamente consciente de algo interno y mental y sólo indirectamente consciente de los objetos externos como los árboles y las hojas.

Un argumento análogo acerca de las pinturas partiría de la premisa de que una pintura puede ser la pintura de un unicornio aunque no haya unicornios. De esto podría concluirse que la pintura es "en primera instancia" una pintura de otra cosa que es real, por ejemplo, de la idea del pintor de un unicornio.

Para ver que tales argumentos son falaces, consideremos el argumento correspondiente que se aplica a las búsquedas: "Ponce de León buscaba la Fuente de la Juventud. Pero no hay tal cosa. Por lo tanto, debió estar buscando algo mental." Esto es simplemente un error. A partir del hecho de que no existe la Fuente de la Juventud, no se sigue que Ponce de León estuviera buscando algo mental. En particular, no buscaba una idea de la Fuente de la Juventud. Él ya tenía la idea. Lo que quería era una Fuente de la Juventud real, no sólo la idea de ella.

El pintor ha realizado un cuadro de un unicornio. El cuadro pintado no es la pintura de una idea de un unicornio. El pintor podría no saber cómo pintar el cuadro de una idea, especialmente si no está familiarizado con el arte conceptual. Puede ser que el pintor tenga una idea de un unicornio e intente captar esa idea en su pintura. Pero esto es tanto como decir que su pintura es una pintura de la misma cosa de la cual su idea es una idea. La pintura no es una pintura de la idea, sino una pintura acerca de lo que es la idea.

Del mismo modo, lo que Eloísa ve ante ella es un árbol, sea o no una alucinación. Es decir, el contenido de su experiencia visual es que se le presenta un árbol, no la idea de un árbol. Quizá la experiencia visual de Eloísa involucra cierta clase de

representación mental del entorno. De ello no se sigue que sea consciente de una representación mental. Si hay una representación mental, puede ser que de lo que sea consciente es de aquello representado en ella; pero entonces la representación mental representa algo en el mundo, no algo en la mente.

Ahora bien, en ocasiones consideramos que alguien percibe algo sólo si esa cosa existe. Así, si no hay un árbol ante ella y Eloísa sufre una alucinación, podríamos describir la situación diciendo o bien que Eloísa ve algo que no está realmente ahí o bien que en realidad no ve nada sino que sólo parece ver algo. No existe una aplicación de "buscar" que corresponda a este segundo uso de "ver", el cual nos permitiría decir que, como no hubo ni hay algo como la Fuente de la Juventud, Ponce de León no estaba realmente buscando nada.

Pero esta ambigüedad en los verbos perceptuales no afecta el argumento que estoy tratando de establecer. A fin de esclarecer esto, utilicemos "ver†" ("ver-daga") para el sentido de "ver" en el cual el objeto visto puede no existir, como cuando Macbeth vio una daga ante él.[2] Y utilicemos "ver*" ("ver-asterisco") para el sentido de ver en el cual sólo las cosas que existen pueden verse. Macbeth vio† una daga, pero no vio* una daga.

El argumento de la ilusión parte de un caso en el cual Eloísa "ve" algo café y verde ante ella, aunque no hay nada café y verde ante ella en el mundo físico externo. De esto, el argumento infiere que lo café y verde que ella ve debe ser interno y mental. Ahora bien, si "ver" es aquí "ver†", ésta es la falacia ya señalada, como la de concluir que Ponce de León buscaba una cosa mental a partir del hecho de que no existe ninguna Fuente de la Juventud en el mundo externo. Por otro lado, si "ver" es aquí "ver*", entonces la premisa del argumento simplemente comete una petición de principio. No se ha ofrecido ninguna razón hasta ahora para la afirmación de que Eloísa ve* algo café y

[2] W. Shakespeare, *Macbeth*, Acto II, Escena I: "¿Es una daga lo que veo, con el puño vuelto hacia mí? Ven, deja que te tome. No puedo tocarte, y aún te sigo viendo. ¿No eres, fatal visión, tan sensible al tacto como a la vista? ¿O eres sólo una daga de la mente, una falsa creación que procede del ardor del cerebro oprimido? [...] Aún te veo y en tu hoja y tu puño gotea ya la sangre, ausente hace unos momentos. No hay tal cosa; es el asunto sangriento mismo que habla así a mis ojos."

verde en este caso. Es cierto que su experiencia perceptual la representa como visualmente presentada con algo café y verde; pero esto es meramente decir que ve† algo café y verde, y no que ve* algo en lo absoluto. (En adelante suprimiré la modificación de los verbos perceptuales marcada con † y *, a menos que sea crucial indicar cuál es el sentido deseado para la discusión.)

En este punto, algunos filósofos (por ejemplo, Jackson 1977) objetarían de la siguiente manera:

> Estás de acuerdo en que hay un sentido en el cual Eloísa ve algo verde y café cuando no hay nada verde y café ante ella en el mundo externo. Eres capaz de negar que esta cosa café y verde sea mental, considerándola inexistente y un objeto meramente intencional. Pero es sin duda más razonable suponer que uno es, en este caso, consciente de algo mental, que suponer que uno es consciente de algo que no existe. ¿Cómo puede haber algo que no existe? La sugerencia misma es una contradicción de los términos, ya que "ser" significa simplemente "existir", de suerte que estás diciendo en realidad que existe algo que no existe (Quine 1948). ¡No hay objetos inexistentes!

Como respuesta, déjenme conceder inmediatamente que no tengo una teoría perfectamente desarrollada de los objetos intencionales. Parsons (1980) ofrece una teoría así, aunque no pretendo expresar una opinión en cuanto al éxito del enfoque de Parsons. De hecho, estoy muy dispuesto a creer que no hay en realidad objetos inexistentes y que el discurso aparente sobre tales objetos debería disolverse de algún modo en el análisis. No me parece que sea mi labor resolver este asunto. Independientemente de cómo se resuelva esta cuestión, será mejor que la teoría resultante termine por coincidir en que Ponce de León estaba buscando algo cuando buscaba la Fuente de la Juventud, aunque no exista ninguna Fuente de la Juventud, y mejor será que la teoría *no* tenga la consecuencia de que Ponce de León estaba buscando algo mental. Si una teoría lógica puede dar cuenta de las búsquedas de cosas que, de hecho, no existen, es de suponerse que también puede permitir un sentido de "ver" de acuerdo con el cual Macbeth pueda ver algo que no existe realmente.

Otra consideración es que la experiencia visual de Eloísa no sólo presenta un árbol. Presenta un árbol visto desde cierto lugar. Varias características con las que se presenta el árbol se presentan como relaciones entre el espectador y el árbol, por ejemplo, las características que el árbol tiene desde aquí. El árbol se presenta como "en frente de" y "ocultando" otros árboles. Se presenta como más frondoso hacia "la derecha". Se presenta como del mismo tamaño "desde aquí" que un árbol más pequeño que está más cercano, lo cual no significa que realmente se vea de igual tamaño, sino sólo que se lo presenta como subtendente aproximadamente al mismo ángulo desde aquí que el árbol más pequeño. Ser presentado como del mismo tamaño desde aquí no es ser presentado como del mismo tamaño, punto.

No pretendo sugerir que la manera en que se presenta visualmente el árbol como desde aquí sea algo que pueda expresarse fácilmente en palabras. En particular, no pretendo sugerir que el árbol pueda, por ende, ser presentado como subtendiente a cierto ángulo visual sólo a una persona que entiende palabras como "subtender" y "ángulo" (como se supone en Peacocke 1983, cap. I). Sólo quiero decir que esta característica de un árbol desde aquí es una característica objetiva del árbol en relación con el aquí, una característica a la cual los perceptores son sensibles y la cual su experiencia visual puede representar de algún modo como que las cosas tienen desde aquí.

Ahora bien, quizá la experiencia visual de Eloísa le presenta incluso un árbol como visto por ella, es decir, como un objeto de su experiencia visual. De ser así, hay después de todo un sentido en el cual la experiencia visual de Eloísa representa algo mental: representa los objetos del mundo como objetos de la experiencia visual. Pero esto no significa que la experiencia visual de Eloísa le revele en modo alguno las propiedades intrínsecas de esa experiencia en virtud de las cuales posee el contenido que posee.

Quiero enfatizar esta observación, porque es muy importante. Eloísa es consciente del árbol como un árbol que ahora está viendo. Así, podemos suponer que es consciente de ciertas características de su experiencia visual que está teniendo. En particular, es consciente de que su experiencia visual tiene la

característica de ser una experiencia de ver un árbol. Esto es ser consciente de una característica intencional de su experiencia; es consciente de que su experiencia posee cierto contenido. Por otro lado, quisiera argumentar que Eloísa no es consciente de esas características intrínsecas de su experiencia en virtud de las cuales posee ese contenido. De hecho, creo que ella no puede acceder de ningún modo a las características intrínsecas de su representación mental que la hacen una representación mental de ver un árbol.

Las cosas son diferentes con las pinturas. En el caso de una pintura, Eloísa puede ser consciente de las características de la pintura que son responsables de que sea una pintura de un unicornio. Esto es, puede prestar atención al patrón de las pinturas sobre el lienzo, en virtud del cual el cuadro representa un unicornio. Pero en el caso de su experiencia visual de un árbol, quiero señalar que ella no es consciente de, por así decirlo, la pintura mental en virtud de la cual su experiencia es una experiencia de ver un árbol. Sólo es consciente de las características intencionales o relacionales de su experiencia, no de sus características no intencionales e intrínsecas.

Algunos teóricos de los datos sensoriales objetarán que Eloísa es, de hecho, consciente de la pintura mental pertinente cuando es consciente de un patrón de colores, porque estos teóricos de los datos sensoriales afirman que el color del cual ella es consciente es interno y mental, y no una propiedad de los objetos externos. Pero esta afirmación acerca de la teoría de los datos sensoriales va en contra de la experiencia visual ordinaria. Cuando Eloísa ve un árbol ante ella, los colores que experimenta son experimentados todos como características del árbol y de sus alrededores. No experimenta ninguno como característica intrínseca de su experiencia. Tampoco experimenta ninguna de las características de otras cosas como características intrínsecas de su experiencia. Y esto es verdadero también en tu caso. No hay nada especial acerca de la experiencia visual de Eloísa. Cuando ves un árbol, no experimentas ninguna de sus características como características intrínsecas de tu experiencia. Contempla un árbol e intenta fijar la atención en las características intrínsecas de tu experiencia visual. Predigo que descubrirás que las únicas características sobre las cuales

fijar la atención ahí serán las características del árbol presentado, incluyendo las características relacionales del árbol "desde aquí".

La idea de los teóricos de los datos sensoriales acerca de nuestra experiencia inmediata del color no es, definitivamente, la idea ingenua; no representa el punto de vista de la percepción ordinaria. La teoría de los datos sensoriales no es resultado de un estudio fenomenológico; es más bien el resultado de un argumento, a saber, del argumento de la ilusión. Pero, como hemos visto, ese argumento o bien es inválido o bien comete una petición de principio.

Es muy importante distinguir lo que se experimenta como características intrínsecas del objeto intencional de la experiencia de las características intrínsecas de la experiencia misma. No es siempre fácil distinguir estas cosas, pero pueden distinguirse. Considera la experiencia de tener un dolor en tu pierna derecha. Es muy tentador confundir las características de lo que experimentas como si ocurriera en tu pierna con las características intrínsecas de tu experiencia. Pero lo que a ti se te presenta como que te ocurre en la pierna es el objeto intencional de tu experiencia; no es la experiencia misma. El contenido de tu experiencia es que hay un problema de cierta clase específica en tu pierna derecha. El objeto intencional de la experiencia es un suceso localizado en tu pierna derecha. La experiencia misma no está localizada en tu pierna derecha. Si la experiencia está en un lugar específico, está en cierta parte de tu cerebro.

Observemos que el contenido de tu experiencia puede no ser verdad con respecto a lo que de hecho está sucediendo. Un disco dislocado en tu espalda puede estar presionando el nervio ciático, haciendo que parezca que tienes un problema en la pierna derecha cuando en realidad no es así. El objeto intencional de tu experiencia de dolor puede no existir. Desde luego, esto no significa que no tengas dolor en la pierna. Sí sientes algo ahí. Pero hay un sentido en el cual lo que sientes en tu pierna es una ilusión o una alucinación.

Es verdad que si Melvin alucina un elefante rosa, el elefante que Melvin ve no existe. Pero el dolor en su pierna, resultante

de la dislocación de un disco en su espalda sin duda existe.[3] El dolor no es un objeto intencional en la misma forma en que lo es el elefante. El dolor en su pierna, causado por la dislocación del disco en su espalda, se parece más a la imagen residual producida por una luz brillante. Si miras una pared blanca, verás la imagen allí. La imagen está en la pared, el dolor está en tu pierna. No hay un punto físico en la pared, no hay un problema físico en tu pierna. La imagen residual existe, el dolor existe. Cuando hablamos de imágenes residuales o de dolores por repercusión, parte de lo que decimos es acerca de nuestra experiencia y parte de lo que decimos es acerca del objeto intencional de esa experiencia. Cuando decimos que el dolor o la imagen residual existen, queremos decir que la experiencia existe. Cuando decimos que la imagen residual está en la pared o que el dolor está en tu pierna, estamos hablando de la localización del objeto intencional de esa experiencia.

3. *Evaluación de la primera objeción*

Estamos ahora en posición de rebatir el primero de los tres argumentos contra el funcionalismo, el cual repetiré ahora:

> Cuando uno presta atención a un dolor en la pierna o a su experiencia de lo rojo de una manzana, es consciente de una cualidad intrínseca de su experiencia, donde una cualidad intrínseca es una cualidad que algo tiene en sí mismo, aparte de sus relaciones con otras cosas. Esta cualidad de la experiencia no puede captarse en una definición funcional, ya que a tal definición le conciernen solamente las relaciones, relaciones entre los estados mentales y la entrada perceptual, relaciones de los estados mentales entre sí y relaciones entre los estados mentales y la salida conductual.

Podemos ver ahora que este argumento falla al confundir una cualidad del objeto intencional de la experiencia con una cualidad de la experiencia misma. Cuando prestas atención a un dolor en tu pierna o a la experiencia de lo rojo de una manzana, estás prestando atención a una cualidad de algo que ocurre en

[3] Estoy en deuda con Sydney Shoemaker por llamarme la atención sobre esto.

tu pierna o a una cualidad de la manzana. Quizá esta cualidad se te presente como una cualidad intrínseca de lo que ocurre en tu pierna o como una cualidad intrínseca de la superficie de la manzana. Pero no se presenta en modo alguno como una cualidad intrínseca de tu experiencia. Y, puesto que no eres consciente del carácter intrínseco de tu experiencia, el hecho de que el funcionalismo se abstraiga del carácter intrínseco de la experiencia no demuestra que deje fuera algo de lo que eres consciente.

Sin duda, hay complicaciones posibles. Supongamos que David se somete a una cirugía cerebral que observa en un espejo. Supongamos que ve ciertas características intrínsecas de la activación de ciertas neuronas en su cerebro, y supongamos que la activación de estas neuronas es la realización de parte de la experiencia que él está teniendo en ese momento. En tal caso, David es consciente de características intrínsecas de su experiencia. Pero esta manera de ser consciente de las características intrínsecas de la experiencia no es incompatible con el funcionalismo. Dada una explicación funcionalista de la percepción que David tiene de los árboles, las mesas y los procesos cerebrales de otras personas, la misma explicación se aplica cuando ocurre que el objeto percibido son los procesos cerebrales del propio David. La conciencia que David tiene de sus propios procesos cerebrales es psicológicamente similar a la conciencia que cualquier otro perceptor visual podría tener de esos mismos procesos cerebrales, incluyendo a perceptores conformados de manera muy distinta de la manera en que David está conformado.

De acuerdo con el funcionalismo, las propiedades psicológicamente pertinentes de un proceso interno son todas propiedades funcionales. La naturaleza intrínseca del proceso es pertinente sólo a medida que es responsable de que el proceso tenga las propiedades funcionales que tiene. He estado considerando la objeción de que, además de la contribución de una experiencia a su función, ciertas características intrínsecas de la experiencia deben ser propiedades psicológicamente pertinentes, ya que éstas son propiedades de las que somos o podemos ser conscientes. La objeción no es sólo que podemos adquirir conciencia de las características intrínsecas de ciertos

procesos mentales en la forma que acabamos de mencionar, es decir, percibiendo en un espejo los procesos físicos subyacentes que realizan esos procesos mentales. Ésta no sería una objeción al funcionalismo. La objeción es más bien que toda o la mayor parte de la experiencia consciente posee aspectos intrínsecos de los cuales somos o podemos ser conscientes de manera tal que estos aspectos de la experiencia son psicológicamente importantes por encima y más allá de la contribución que hagan a la función.

Desde luego, decir que estos aspectos son psicológicamente importantes no equivale a afirmar que son o deberían ser importantes para la ciencia de la psicología. Más bien, se supone que son psicológicamente importantes en el sentido de que resultan mentalmente importantes, sea o no este aspecto de la experiencia susceptible de comprensión científica. La objeción es que cualquier explicación de nuestra vida mental que no considere estas propiedades intrínsecas como mentales o psicológicas deja fuera un aspecto crucial de nuestra experiencia.

Mi respuesta a esta objeción es que no se la puede defender sin confundir las características intrínsecas del objeto intencional de la experiencia con las características intrínsecas de la experiencia. Aparte de esta confusión, no hay razón para pensar que somos alguna vez conscientes de las características intrínsecas pertinentes de nuestras experiencias.

Hay otras formas en las que se puede ser consciente de las características intrínsecas de nuestra experiencia, sin que esto arroje ninguna duda sobre el funcionalismo. Por ejemplo, se puede ser consciente de las características intrínsecas de la experiencia sin ser consciente de ellas como características intrínsecas de la experiencia, tal como Ortcutt puede ser consciente de un hombre, quien resulta ser un espía, sin ser consciente del hombre como un espía. Cuando Eloísa ve un árbol, es consciente de su experiencia perceptual como una experiencia con cierto contenido intencional. Supongamos que su experiencia se realiza en un suceso físico particular y que ciertas características intrínsecas del suceso son, en este caso, responsables de ciertas características intencionales de la experiencia de Eloísa. Quizá hay entonces un sentido en el cual Eloísa es

consciente de este proceso físico y consciente de esas características intrínsecas, aunque no es consciente de ellas como las características intrínsecas que son.

Aun si esto es así, no constituye una objeción al funcionalismo. Las características intrínsecas de las cuales Eloísa es consciente no son más importantes psicológicamente que la propiedad de ser un espía a la percepción que tiene Ortcutt de un hombre que resulta ser un espía. El caso no ofrece razones para pensar que existe una diferencia psicológicamente importante entre la experiencia de Eloísa y la experiencia de cualquier duplicado funcional de Eloísa que está compuesto por material diferente del que compone a Eloísa.

Asimismo, si Eloísa se somete a la clase de educación que recomienda Paul Churchland (1985), de suerte que piensa automáticamente en los aspectos intencionales de su experiencia en términos de sus causas neurofisiológicas, entonces puede ser consciente de las características intrínsecas de su experiencia como las características mismas que son. Pero tampoco esto sería una objeción al funcionalismo, ya que no ofrece motivo alguno para pensar que hay una diferencia psicológica entre Eloísa después de esa educación y un robot que es el duplicado funcional de Eloísa y que ha recibido una educación similar (Shoemaker 1985). El duplicado piensa ahora erróneamente en ciertos aspectos de su experiencia como si fueran características de ciertos procesos neurológicos, y lo hace erróneamente porque los procesos pertinentes en el duplicado no son en modo alguno procesos neurológicos.

Observemos, por cierto, que no estoy ofreciendo ningún tipo de argumento positivo en el sentido de que Eloísa y su duplicado deben tener experiencias que sean psicológicamente semejantes en todos sus aspectos. Sólo estoy haciendo la observación de que los casos que acabamos de considerar son compatibles con la afirmación funcionalista de que sus experiencias son similares.

Las objeciones al funcionalismo que estoy considerando en este trabajo afirman que ciertas propiedades intrínsecas de la experiencia conforman de tal suerte la experiencia que cualquier experiencia con propiedades intrínsecas diferentes tendría un carácter psicológico distinto. Lo que he argumentado

hasta aquí es que esta objeción no se sostiene mediante una simple inspección de nuestra experiencia.

4. *Percepción y entendimiento*

Ahora bien, pasemos a la segunda objeción, que voy a repetir:

> Una persona ciega de nacimiento podría saber todo acerca de los hechos físicos y funcionales de la percepción del color, sin saber cómo es ver algo rojo. Por lo tanto, cómo es ver algo rojo no puede explicarse en términos puramente funcionales.

A fin de contestar esta objeción, debo hablar de la teoría funcionalista del contenido de las representaciones mentales y, más particularmente, de la teoría funcionalista de los conceptos. Debo hacer esto porque saber cómo es ver algo rojo equivale a ser capaz de representarse a sí mismo algo como rojo. Uno puede representarse esto a sí mismo sólo si tiene el concepto pertinente de qué es que algo sea rojo. La persona ciega carece del concepto pleno de lo rojo que posee una persona que ve; así, la persona ciega no puede representarse plenamente qué es para una persona que ve ver algo rojo. Por lo tanto, no puede decirse que la persona ciega sepa cómo es ver algo rojo.

Un tipo de explicación funcionalista de la representación mental supone que las representaciones mentales se construyen a partir de conceptos, donde el contenido de una representación se determina por los conceptos que contiene y la forma en que estos conceptos se integran para formar esa representación (Harman 1987). Desde este punto de vista, qué es tener un concepto dado se determina funcionalmente. Alguien tiene el concepto apropiado de algo como rojo si y sólo si la persona tiene a su disposición un concepto que funcione de manera adecuada. El funcionamiento pertinente puede involucrar conexiones con el uso de otros conceptos, conexiones con la entrada perceptual, y/o conexiones con la salida conductual. En este caso, las conexiones con las entradas perceptuales son cruciales. Para que el concepto funcione de manera tal que la persona tenga el concepto pleno de algo como rojo, la persona debe estar dispuesta a formarse representaciones que involucren ese concepto como la consecuencia natural e inmediata

de ver algo rojo. Puesto que la persona ciega carece de un concepto de este tipo, la persona ciega carece del concepto pleno de algo como rojo. Por lo tanto, la persona ciega no sabe cómo es ver algo rojo.

No es fácil especificar con precisión la relación funcional pertinente. Alguien que se queda ciego en una etapa posterior de su vida conservará normalmente el concepto pertinente de algo como rojo. Tal persona posee un concepto que podría usar al formarse esas representaciones visuales inmediatas, salvo por el problema que en su caso interfiere con la percepción visual normal. Así, la relación funcional correcta se sostiene para tal persona. Estoy suponiendo que la persona ciega de nacimiento carece de ese concepto; es decir, tal persona no tiene un concepto de algo como rojo al que podría recurrir de inmediato en las representaciones visuales del entorno si de pronto adquiriera la vista.

Estamos ahora en posición de evaluar la afirmación de que la persona ciega de nacimiento podría conocer todos los hechos físicos y funcionales sobre la percepción del color sin saber cómo es ver algo rojo. Sostengo que hay un hecho funcional importante sobre la percepción de color que la persona ciega no puede conocer, a saber, que existe un concepto *R* tal que, cuando un perceptor normal ve algo rojo en condiciones de iluminación buenas, el perceptor tiene una experiencia visual con una estructura representacional que contiene este concepto *R*. La persona ciega de nacimiento desconoce este hecho, porque, para conocerlo, necesita poder representárselo, lo cual exige tener los conceptos pertinentes. Un concepto clave que se requiere para representarse ese hecho es el concepto de algo como rojo, porque el hecho en cuestión es un hecho sobre lo que sucede cuando un perceptor normal ve algo rojo. Puesto que la persona ciega de nacimiento no tiene el concepto pleno de algo como rojo, no puede comprender plenamente ese hecho y así no puede conocer ese hecho.

La persona ciega podría saber algo semejante a esto, por ejemplo, que hay un concepto *R* tal que, cuando un perceptor normal ve algo que refleja la luz de tal y cual frecuencia, el perceptor tiene una experiencia visual con una estructura re-

presentacional que contiene este concepto *R*. Pero esto es saber algo diferente.

El ciego de nacimiento no puede saber cómo es ver algo rojo porque no comprende plenamente qué es que algo sea rojo, esto es, porque carece del concepto pleno de algo como rojo. Por lo tanto, al contrario de lo que se supone en la segunda objeción, la persona ciega de nacimiento no conoce todos los hechos funcionales, ya que no sabe cómo funciona el concepto *R* con respecto a la percepción de las cosas que son rojas.

Esta respuesta a la segunda objeción recurre a un funcionalismo que hace referencia a las funciones de los conceptos, no sólo a las funciones de los estados mentales en general. Hay otras versiones del funcionalismo que intentan arreglárselas refiriéndose a las funciones de los estados mentales en general, sin recurrir a los conceptos. Algunas de estas versiones identifican los contenidos de dichos estados con conjuntos de mundos posibles (o mundos posibles centrados). Estas versiones del funcionalismo no pueden responder a la objeción en la forma en que yo lo he hecho. No me resulta claro si es posible una respuesta satisfactoria por parte de tales teorías. Por ejemplo, Lewis (1983) se ve obligado a decir que, aunque un ciego de nacimiento carece de ciertas habilidades, por ejemplo, de la capacidad de reconocer los objetos rojos con sólo mirarlos de la manera en que pueden hacerlo las personas que ven, el ciego no carece de ninguna información sobre la percepción visual. No estoy conforme con esta respuesta, ya que es claramente falso decir que el ciego de nacimiento no carece de ninguna información.

5. *El espectro invertido*

Pasaré ahora a la tercera objeción al funcionalismo, que reproduzco:

> Es concebible que dos personas tengan sistemas visuales que funcionan de manera similar, a pesar del hecho de que las cosas que le parecen rojas a una persona le parecen verdes a la otra, las cosas que le parecen anaranjadas a la primera persona le parecen azules a la segunda, y así sucesivamente. Este tipo de inversión del espectro en la manera en que se ven las cosas es posible, pero no puede dársele una descripción puramente funcional, pues, por hipótesis,

> no existen diferencias funcionales entre las personas en cuestión. Puesto que la manera en que se ven las cosas para una persona es un aspecto de la vida mental de esa persona, esto significa que hay un aspecto importante de la vida mental de una persona que no puede explicarse en términos puramente funcionales.

Para discutir esta objeción, necesito decir un poco más sobre cómo funcionan los estados perceptuales. En particular, debo decir algo acerca de la manera en que funcionan los estados perceptuales en relación con las creencias.

La experiencia perceptual representa un entorno particular del perceptor. Normalmente, un perceptor utiliza esta representación como su representación del entorno. Es decir, el perceptor la utiliza a fin de sortear el mobiliario. O también en otras palabras, esta representación se usa como la creencia que tiene el perceptor acerca del entorno. Esta forma de usar las representaciones perceptuales es el caso normal, aunque hay excepciones cuando un perceptor inhibe su tendencia natural y se abstiene de utilizar una representación perceptual (o ciertos aspectos de esa representación) como guía para su entorno, como una creencia acerca de lo que le rodea. El contenido de la representación perceptual se define funcionalmente, en parte, por la manera en que esta representación surge normalmente en la percepción y, en parte, por la manera en que la representación se utiliza normalmente para guiar las acciones (Armstrong 1961, 1968; Dennett 1969; Harman 1973).

La objeción nos lleva a considerar a dos personas, llamémoslas Alicia y Alfredo, con sistemas visuales que funcionan de un modo similar, pero cuyos espectros se hallan invertidos uno respecto del otro. Las cosas que a Alicia le parecen rojas, a Alfredo le parecen verdes, las cosas que a Alicia le parecen azules, le parecen anaranjadas a Alfredo, y así sucesivamente. Hemos de imaginar que esta diferencia entre Alicia y Alfredo no se refleja en lo absoluto en su comportamiento. Ambos llaman "rojas" a las fresas maduras y "verde" a la hierba, y lo hacen con la misma naturalidad con que lo hacen los perceptores normales que han aprendido español de la manera habitual.

Consideremos lo que esto significa para Alicia en un caso normal de percepción. Ella ve una fresa madura. El procesa-

miento perceptual tiene por resultado una representación perceptual de esa fresa, incluida una representación de su color. Alicia utiliza esta representación como su guía para el entorno, es decir, como su creencia acerca de la fresa y, en particular, como su creencia acerca de su color. Expresa su creencia sobre el color de la fresa utilizando las palabras "es roja". Lo mismo ocurre con Alfredo. Su percepción de la fresa tiene por resultado una representación perceptual del color de la fresa que utiliza como su creencia sobre el color y que expresa con las mismas palabras, "es roja".

Ahora bien, en el caso normal de percepción, no puede haber distinción alguna entre cómo se ven las cosas y cómo se cree que son, ya que la manera como se ven las cosas está dada por el contenido de nuestra representación perceptual y, en el caso normal, nuestra representación perceptual se usa como nuestra creencia acerca del entorno. La hipótesis de la objeción del espectro invertido es que la fresa aparece como siendo de colores diferentes para Alicia y Alfredo. Puesto que se supone que todo funciona en ellos de modo normal, se sigue que deben tener creencias diferentes con respecto al color de la fresa. Si tuvieran las mismas creencias cuando tienen representaciones perceptuales que difieren en cuanto a su contenido, entonces por lo menos uno de ellos tendría una representación perceptual que no estaría funcionando como su creencia acerca del color de la fresa, lo que equivale a decir que no estaría funcionando según lo que suponemos que es la manera normal.

Una consecuencia adicional de la hipótesis del espectro invertido es que, dado que en el caso normal Alicia y Alfredo expresan sus creencias sobre el color de las fresas y de la hierba diciendo "es roja" y "es verde", deben significar algo distinto con sus palabras de colores. Por "rojo", Alfredo significa la manera en que las fresas maduras le parecen a él. Puesto que ésta es la manera en que la hierba se le presenta a Alicia, lo que Alfredo significa con "rojo" es lo que ella significa con "verde".

Es importante advertir que éstas son realmente consecuencias de la hipótesis del espectro invertido. Si Alicia y Alfredo significaran lo mismo con sus términos de colores, entonces o bien (a) uno de ellos no estaría usando estas palabras para expresar sus creencias acerca del color, o bien (b) uno de ellos

no estaría usando sus representaciones perceptuales del color como sus creencias sobre el color. En cualquiera de estos casos, habría una falla del funcionamiento normal, contrario a lo que plantea la hipótesis de la objeción del espectro invertido.

·De acuerdo con el funcionalismo, si Alicia y Alfredo emplean palabras en el mismo sentido con respecto a las mismas cosas, entonces significan lo mismo con esas palabras (suponiendo también que son miembros de la misma comunidad lingüística y que toman sus palabras de un lenguaje común). Pero esto es simplemente sentido común. Supongamos que Alicia y Humphrey son ambos miembros de una misma comunidad lingüística que utilizan las palabras de la misma manera, etc. Alicia es un ser humano ordinario, mientras que Humphrey es un robot humanoide fabricado con un material totalmente distinto del de Alicia. El sentido común atribuiría los mismos significados a las palabras de Humphrey que a las de Alicia, dado que utilizan las palabras del mismo modo. Se requiere cierto tipo de argumento filosófico para argüir lo contrario. Los defensores de la objeción del espectro invertido no han presentado ningún argumento semejante.

Shoemaker (1982) ofrece una versión diferente de la objeción del espectro invertido. Nos hace considerar a una sola persona, llamémosla Harry, en dos momentos distintos: en una etapa inicial de percepción normal de color y en una etapa posterior, después de haber sufrido una inversión del espectro sumamente notoria (quizá como resultado del tipo de operación cerebral descrita en Lycan 1973, en la cual se intercambian los nervios de suerte que las cosas rojas tienen ahora las consecuencias perceptuales que solían tener las cosas verdes, etc.) y de haber finalmente adaptado por completo sus respuestas hasta restablecer su funcionamiento normal. Shoemaker concuerda en que Harry posee ahora las mismas creencias sobre el color que antes y significa las mismas cosas con sus palabras sobre el color, y también concuerda en que hay un sentido en el cual las fresas le parecen ahora a Harry iguales a como las veía antes de la inversión del espectro. Sin embargo, Shoemaker considera evidente que hay otro sentido de "verse" según el cual bien puede ser verdad que las cosas no se vean iguales a como se veían antes, de modo que en este segundo sentido de "verse",

las cosas rojas se ven de la misma manera en que solían verse las cosas verdes.

En otras palabras, Shoemaker cree que es evidente que puede haber una diferencia psicológicamente pertinente entre la clase de experiencia que tenía Harry al mirar una fresa madura en la etapa inicial y la experiencia que tiene al mirar una fresa madura en la etapa final (después de haberse adaptado por completo a su operación). Es decir, piensa que es evidente que puede haber una diferencia psicológicamente pertinente entre estas experiencias aun cuando no exista ninguna diferencia funcional y ninguna diferencia en el contenido de las experiencias.

Ahora bien, esto puede parecerle evidente a alguien que incurra en la falacia de los datos sensoriales, la cual sostiene que la conciencia que uno tiene del color de una fresa está mediada por la conciencia de una característica intrínseca de la representación perceptual. Pero, ¿por qué debería cualquier otra persona estar de acuerdo con esto? Dos experiencias perceptuales con el mismo contenido intencional deben ser psicológicamente iguales. En particular, no puede haber algo de lo que uno sea consciente al tener una experiencia y de lo que uno no sea consciente al tener la otra, ya que el contenido intencional de una experiencia comprende todo aquello de lo que uno es consciente al tener esa experiencia.

Mi sugerencia es que la hipótesis del espectro invertido de Shoemaker sólo le parecerá evidente a alguien que *parta* del supuesto previo de que las personas tienen una conciencia inmediata y directa de las características intrínsecas de su experiencia, incluidas las características intrínsecas que funcionan para representar el color. Tal persona puede entonces pasar a suponer que la característica intrínseca de la experiencia que representa lo rojo para Alicia es la característica intrínseca de la experiencia que representa lo verde para Alfredo, y así sucesivamente. Este supuesto previo es precisamente la idea detrás de la primera objeción, de la cual he argumentado que se opone a la experiencia ordinaria y que sólo puede ser defendida confundiendo las cualidades de los objetos intencionales de la experiencia con las cualidades de la experiencia misma. Luego, la

hipótesis del espectro invertido de Shoemaker no ofrece ningún argumento independiente en contra del funcionalismo.[4]

6. *Conclusión*

Para resumir en pocas palabras, he descrito y contestado tres objeciones relacionadas que se le plantean al funcionalismo. La primera afirma que tenemos una conciencia directa de las características intrínsecas de nuestra experiencia y argumenta que no hay forma de dar cuenta de esta conciencia desde un punto de vista funcional. A esto respondo que, cuando distinguimos claramente las propiedades del objeto de la experiencia de las propiedades de la experiencia, vemos que no tenemos conciencia de las características intrínsecas pertinentes de la experiencia. La segunda objeción afirma que una persona ciega de nacimiento puede saberlo todo acerca del papel funcional de la experiencia visual sin saber cómo es ver algo rojo. A esto respondo que el ciego no lo sabe todo acerca del papel funcional de la experiencia visual; en particular, la persona ciega no sabe cómo funciona esa experiencia en relación con la percepción de los objetos rojos. La tercera objeción afirma que el funcionalismo no puede dar cuenta de la posibilidad de un espectro invertido. A esto respondo que alguien que sufre de la clase pertinente de espectro invertido tendría que tener creencias sobre los colores de las cosas que fueran distintas de las creencias que los otros tienen y tendría que significar algo distinto con sus términos de colores, pese a ser un perceptor de colores funcionalmente normal, que clasifica las cosas según sus colores en exactamente la misma manera en que los demás lo hacen y que utiliza la terminología de los colores del mismo modo en que lo hacen los otros. El rechazo del funcionalismo a esta posibilidad es de sentido común y, ciertamente, no es tan absolutamente inverosímil o contraintuitivo que haga que estos casos representen una objeción al funcionalismo. Por otro lado, imaginar que podrían presentarse casos pertinentes de espectro

[4] Debo decir que Shoemaker mismo no ofrece su caso como una objeción a lo que él llama funcionalismo. Afirma que su versión del funcionalismo es compatible con su caso. Pero yo estoy considerando una versión del funcionalismo definida de tal forma que se vuelve incompatible con ese caso.

invertido sin una inversión de las creencias y los significados es caer otra vez en la primera objeción, y no ofrecer una consideración adicional en contra del funcionalismo.

[*Traducción de Ana Isabel Stellino*]

BIBLIOGRAFÍA

Anscombe, G.E.M., 1981, *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers*, University of Minnesota Press, Minneápolis.

———, 1965, "The Intentionality of Sensation: A Grammatical Feature", en Butler 1965, pp. 158–180. (Reproducido en Anscombe 1981, vol. II, pp. 3–20.)

Armstrong, D.M., 1968, *The Materialist Theory of Mind*, Routledge and Kegan Paul, Londres.

———, 1962, *Bodily Sensations*, Routledge and Kegan Paul, Londres.

———, 1961, *Perception and the Physical World*, Routledge and Kegan Paul, Londres. [Versión en castellano: *La percepción y el mundo físico*, trad. P. García Ferrero, Tecnos, Madrid, 1966.]

Block, N., 1986, "Advertisement for a Semantics for Psychology", *Midwest Studies in Philosophy*, vol. 10, pp. 615–678.

Butler, R.J. (comp.), 1965, *Analytical Philosophy. Second Series*, Blackwell, Oxford.

Churchland, P., 1985, "Reduction, Qualia, and the Direct Introspection of Mental States", *Journal of Philosophy*, vol. 82, pp. 8–28.

Dennett, D.C., 1987, *The Intentional Stance*, MIT Press, Cambridge, Mass. [Versión en castellano: *La actitud intencional*, trad. D. Zadunaisky, Gedisa, Barcelona, 1991.]

———, 1969, *Content and Consciousness*, Routledge and Kegan Paul, Londres. [Versión en castellano: *Contenido y conciencia*, trad. J.M. Lebrón, Gedisa, Barcelona, 1996.]

Harman, G., 1988, "Wide Functionalism", en S. Schiffer y S. Steele (comps.), *Cognition and Representation*, Westview Press, Boulder, pp. 11–20.

———, 1987, "(Nonsolipsistic) Conceptual Role Semantics", en E. LePore (comp.), *New Directions in Semantics*, Academic Press, Londres, pp. 55–81.

———, 1973, *Thought*, Princeton University Press, Princeton.

Jackson, F., 1986, "What Mary Didn't Know", *Journal of Philosophy*, vol. 83, pp. 291–295. [Incluido en esta antología, pp. 111–118.]
———, 1982, "Epiphenomenal Qualia", *Philosophical Quarterly*, vol. 32, no. 127, pp. 127–136. [Incluido en esta antología, pp. 95–110.]
———, 1977, *Perception: A Representative Theory*, Cambridge University Press, Cambridge.
Lewis, D.K., 1983, "Postcript to 'Mad Pain and Martian Pain' ", *Philosophical Papers*, vol. 1, Oxford University Press, Nueva York, pp. 130–132.
Lycan, W.G., 1973, "Inverted Spectrum", *Ratio*, vol. 15, pp. 315–319.
Nagel, T., 1970, "Armstrong on the Mind", *Philosophical Review*, vol. 79, no. 3, pp. 394–403. (Reproducido en N. Block (comp.), *Readings in the Philosophy of Psychology*, vol. 1, Harvard University Press, Cambridge, Mass., pp. 200–206.)
———, 1974, "What Is It Like to Be a Bat?" *Philosophical Review* vol. 83, pp. 435–450. [Incluido en esta antología, pp. 45–63.]
Parsons, Terence, 1980, *Nonexistent Objects*, Yale University Press, New Haven.
Peacocke, C., 1983, *Sense and Content*, Oxford University Press, Oxford.
Pitcher, G., 1971, *A Theory of Perception*, Princeton University Press, Princeton.
Quine, W.V.O., 1960, *Word and Object*, MIT Press, Cambridge, Mass. [Versión en castellano: *Palabra y objeto*, trad. M. Sacristán, Labor, Barcelona, 1968.]
———, 1948, "On What There Is", *Review of Metaphysics*. (Reproducido en *From a Logical Point of View*, Harvard University Press, Cambridge, Massachusetts, 1953, pp. 1–19.)
Rumelhart, D.E. y J.L. McClelland, 1986, *Parallel Distributed Processing*, MIT Press, Cambridge, Mass, 2 vols. [Versión en castellano: *Introducción al procesamiento distribuido en paralelo*, trad. J.C. Gómez Crespo *et al.*, Alianza, Madrid, 1992.]
Shoemaker, S., 1982, "The Inverted Spectrum", *Journal of Philosophy*, vol. 79, pp. 357–381.
———, 1985, "Churchland on Reduction, Qualia, and Introspection", *PSA 1984*, vol. 2 (Philosophy of Science Association), pp. 799–809.
Thagard, P.T., 1986, "Parallel Computation and the Mind-Body Problem", *Cognitive Science*, vol. 10, pp. 301–318.

LA TIERRA INVERTIDA

NED BLOCK

Este artículo nació como respuesta a "La cualidad intrínseca de la experiencia" de Gilbert Harman, que aparece en este volumen, y conserva el mismo formato aunque su objetivo es demostrar que hay un argumento de "inversión" en favor del realismo de los *qualia* y contra el funcionalismo, el cual es mejor que el argumento tradicional sobre el espectro invertido.[1] (El realismo de los *qualia*, en el sentido en que utilizaré el término, es la concepción de que nuestra experiencia posee rasgos mentales intrínsecos.) Quienes no hayan leído el escrito de Harman tal vez prefieran saltarse la última sección de este capítulo.

1. *La falacia de la intencionalización de los* qualia

Como primera aproximación, la hipótesis del espectro invertido es que las cosas que, según coincidimos, son rojas tú las ves de la misma forma en que yo veo las cosas que, según coincidimos, son verdes (y somos funcionalmente idénticos). Hay

[1] La respuesta fue presentada en el Coloquio de Chapel Hill, el 17 de octubre de 1987. Una versión posterior de este artículo fue presentada en las universidades de Oxford, Cambridge y Edimburgo, en Birbeck College y King's College de Londres, y en un encuentro del Grupo Anglofrancés de Filosofía. Agradezco el apoyo del Consejo Estadounidense de Sociedades de Estudios Superiores y de la Fundación Nacional de las Ciencias (DIR88 12559). También estoy en deuda con Martin Davies y Christopher Peacocke por sus comentarios a una versión previa, y agradezco las útiles discusiones acerca de estos temas que he sostenido con Davies y con Sam Guttenplan.

un argumento sencillo que va de la posibilidad de un espectro invertido a la falsedad del funcionalismo: si dos estados mentales diferentes pueden desempeñar exactamente el mismo papel funcional, hay entonces un aspecto de la mentalidad (el aspecto "cualitativo") que elude ser caracterizado en términos de un papel funcional. En términos de la versión maquinicista del funcionalismo: aun si fuésemos computadoras, si tú y yo pudiésemos ser exactamente iguales computacionalmente aunque mentalmente distintos (en lo que respecta a cómo es ver algo rojo), entonces lo mental rebasaría lo computacional.

La "respuesta de limitación"[2] al espectro invertido consistiría en abandonar el funcionalismo como teoría de la experiencia (o por lo menos de su aspecto cualitativo), conservándolo como teoría del aspecto cognoscitivo de la mente. Soy partidario de esta forma de abordar el problema, pero no la desarrollaré aquí. Este escrito está dirigido contra el funcionalista extremo, el cual insiste en ver el funcionalismo como una teoría de toda la mente.[3] El inconveniente de la respuesta de limitación estriba en que, como puede argumentarse, compromete a sus defensores con la posibilidad de un "zombi", un ser que es como nosotros en cuanto a estados cognoscitivos, pero que carece por completo de *qualia*.[4]

Anteriormente expuse una primera aproximación a la hipótesis del espectro invertido. Pero una formulación apropiada de la misma debe recurrir a una distinción entre dos tipos de contenidos de la experiencia, uno de los cuales es una cuestión de la manera en que la experiencia representa al mundo, mientras que la otra es una cuestión de "cómo es" (según la expresión de Tom Nagel) tener esa experiencia. El primero ha sido denominado contenido intencional o representacional; el segundo,

[2] *Cfr.* Haugeland 1978.

[3] En Shoemaker 1975 [incluido en esta antología, pp. 183–212], Shoemaker muestra cómo el funcionalismo puede ofrecer condiciones de identidad para los *qualia* aunque se abandone el intento por caracterizar estados cualitativos particulares. La técnica de Shoemaker asume, no obstante, que los *qualia* ausentes son imposibles, supuesto con el que estoy en desacuerdo.

[4] Para los argumentos en contra de la posibilidad de este zombi, véase Shoemaker 1975, McGinn 1983; y Searle 1992. Contesto a Shoemaker en Block 1980c. Shoemaker responde en Shoemaker 1981.

contenido cualitativo o de sensaciones.[5] En términos de esta distinción, la hipótesis del espectro invertido es la siguiente: cuando tú y yo tenemos experiencias que poseen el contenido intencional de *verse rojo*, el contenido cualitativo que tienes es el mismo que el contenido cualitativo que yo tengo cuando mi experiencia posee el contenido intencional de *verse verde* (y somos funcionalmente idénticos). Este artículo se ocupará de esta versión, bastante drástica, de la afirmación de que existe un hueco entre el contenido intencional y el contenido cualitativo de la experiencia. Pero el énfasis que se pone aquí y en la literatura sobre este caso drástico no debe hacernos olvidar que si la teoría funcionalista de los *qualia* fuera correcta, también excluiría diferencias cualitativas menos sistemáticas entre personas funcionalmente idénticas, diferencias cuya formulación en hipótesis, aunque difícil de desarrollar en detalle, es también menos vulnerable a los numerosos abusos que se han hecho de la hipótesis del espectro invertido.

Si ambos vemos roja la sangre, entonces, en el sentido *intencional* de "verse igual", ambos vemos igual la sangre. (Esto es, la vemos igual con respecto al color. En adelante dejaré de lado esta salvedad, y también, por razones de simplicidad, haré caso omiso del hecho de que hay diferentes tonos de rojo.) El sentido *cualitativo* de "verse igual" puede definirse recurriendo a nociones semejantes a "cómo es" o, como alternativa, recurriendo directamente a la propia hipótesis del espectro invertido. Si tu espectro está invertido con respecto al mío, entonces tú verás las cosas rojas de la misma manera —en el sentido cualitativo— en que yo veo las cosas verdes.

Como señala Shoemaker, resulta fácil pasar de estos sentidos de "verse" [o de "ver"] a los diversos tipos de contenido. Si

[5] La terminología sobre lo intencional y lo cualitativo proviene del artículo de Shoemaker (1982). Este ensayo se reproduce, con un nuevo postscriptum y junto con otros escritos de Shoemaker sobre el mismo tema (entre ellos "Functionalism and Qualia" [incluido en esta antología, pp. 183–212]) en Shoemaker 1984. La terminología sobre lo representacional y lo sensorial fue tomada de Peacocke 1983, caps. 1 y 2. Peacocke utiliza "cómo es" de manera que incluye tanto el contenido representacional como el sensorial. La distinción que establece Frank Jackson entre los sentidos comparativos y fenoménicos de "se ve" se reduce en gran medida a lo mismo. Véase el capítulo 2 de Jackson 1977.

vemos la sangre de la misma manera en el sentido intencional, entonces los contenidos intencionales de nuestras experiencias son los mismos. Si, en el sentido cualitativo, tú ves la sangre de la misma forma en que yo veo la hierba, entonces los contenidos cualitativos de nuestras experiencias son iguales.[6]

Ahora que hemos introducido la distinción entre lo intencional y lo cualitativo, puedo corregir un error vital que han cometido ambos bandos del debate sobre el espectro invertido. Mi observación es que si es posible un espectro invertido, entonces los contenidos de la experiencia **que pueden expresarse en el lenguaje público** (por ejemplo, *verse rojo*) no son contenidos cualitativos, sino más bien contenidos intencionales. **Pues supongamos que la inversión de espectros es abundante**: no habría entonces un espectro de la gran mayoría a pesar de que exista una amplia similitud funcional en los aspectos pertinentes. ¿Cómo podría justificar la afirmación de que veo las cosas rojas como rojas, pero que tú no? ¿Cómo podría aseverar que el contenido cualitativo de mi experiencia como de cosas rojas es *verse rojas*, mientras que el de la tuya es, digamos, *verse azules*? Cualquier argumento que pudiera usar en contra tuya, podría ser utilizado igualmente por ti en contra mía. Y no servirá de nada decir que ambos tenemos razón, cada uno con su propio sentido de "rojo". "Rojo" es una palabra del lenguaje público,

[6] Peacocke 1983 utiliza varios ejemplos curiosos para defender la distinción entre lo representacional y lo sensorial sin comprometerse con el espectro invertido. Por ejemplo, argumenta que al mirar una habitación con un solo ojo, se tiene una experiencia con un contenido representacional que es igual al de la experiencia que uno tiene al mirar con los dos ojos (los mismos objetos se representan separados por la mismas distancias), pero los contenidos sensoriales difieren. Si está en lo correcto, existe entonces una distancia mucho mayor entre, por una parte, la distinción entre lo representativo y lo sensorial y, por otra, el espectro invertido, que yo supongo. Eludiré la cuestión individuando los contenidos intencionales **funcionalmente**. Así, los contenidos intencionales monoculares y los binoculares serán distintos porque funcionan de manera diferente. Por ejemplo, puesto que somos plenamente conscientes de la diferencia y podemos recurrir a ella en nuestros reportes verbales de nuestras experiencias, la experiencia binocular puede dar lugar a reportes distintos de los producidos por la experiencia monocular. Véase la reseña de Michael Tye al libro de Peacocke en Tye 1985, para una discusión de otro de los ejemplos de Peacocke desde el mismo punto de vista.

no un término que signifique una cosa en tus labios y otra en mis labios.

Puesto que recurriré a esta cuestión más adelante, será útil darle un nombre. De manera tendenciosa, describiré como la falacia de la intencionalización de los *qualia* el supuesto de que los contenidos de la experiencia que pueden ser expresados en el lenguaje público, como *verse rojo*, son contenidos cualitativos.[7]

Obsérvese que los contenidos intencionales de la experiencia de los colores deben basarse en las referencias; estos contenidos son "californianos", no neofregeanos. La razón de ello estriba en que los "modos de presentación" de las experiencias de los colores son contenidos cualitativos, y es precisamente con respecto a estos contenidos cualitativos que pueden diferir nuestras experiencias como de cosas rojas, nuestros contenidos de *verse rojo*. Lo que confiere a una experiencia el contenido de *verse rojo* es su relación con las cosas rojas, no su contenido cualitativo. Una manera de expresar la cuestión es decir que cada uno de nosotros puede tener su propio "concepto" de la apariencia del rojo, donde un concepto constituye un contenido cualitativo, pero que tales diferencias en el "concepto" de rojo no conllevan diferencia alguna en si algo se ve rojo para nosotros. El que algo se vea rojo para nosotros depende de si estamos teniendo una experiencia como de algo rojo, no de nuestros distintos "conceptos" de rojo. (Recordemos que estoy haciendo caso omiso de los diversos tonos de un mismo color.)[8]

[7] Una salvedad: resulta útil tener ciertos términos para referir a ese contenido de la experiencia que es mental sin ser intencional. He utilizado "cómo es" y "contenido cualitativo" de esta manera, y por eso no deseo considerar estos términos como si se refirieran a los contenidos intencionales.

[8] La idea que he estado desarrollando sí supone que a menudo los objetos tienen colores, por ejemplo, el libro que estoy mirando ahora es azul. No supone que las preguntas acerca del color de las cosas siempre tienen una respuesta determinada, ni tampoco niega que el color de un objeto sea relativo a todo género de cosas como, por ejemplo, la posición de quien lo percibe. Una mancha de color en una tira cómica puede ser anaranjada desde cierta distancia, pero también roja y amarilla de cerca. Si esto es objetivismo, me alegra ser objetivista. Véase Hardin 1988, para un argumento poderoso contra el objetivismo (y el subjetivismo).

Quiero hacer dos observaciones acerca del argumento contra la intencionalización de los *qualia*. En primer lugar, el argumento no está realmente dirigido contra los escépticos de los *qualia*, como Harman y Dennett, ya que éstos no aceptan la premisa del espectro invertido. Además, creen ya en la conclusión de las formulaciones del argumento, a saber, que *verse rojo* es un contenido intencional.[9] La segunda observación consiste en que aunque espero que los realistas de los *qualia* se convenzan del argumento, no pretendo internarme en el asunto detalladamente en este lugar. Mencioné antes que se podía eludir la cuestión suponiendo que "rojo" y, por lo tanto, "verse rojo", eran términos ambiguos, definidos de manera privada por cada uno de nosotros según su propio contenido cualitativo. Desprecié este supuesto, diciendo que "rojo" era una palabra unívoca del lenguaje público. Pero, desde luego, no es preciso que el asunto termine aquí. Y hay otros ardides más refinados de los que se podría valer un defensor decidido de la definición cualitativa de "rojo". No profundizaré en el asunto porque, dado que mi artículo está dirigido contra los escépticos de los *qualia*, una acción desde la retaguardia contra otros miembros de mi equipo no sería una buena estrategia. Los realistas de los *qualia* que se sienten tentados a intencionalizar los *qualia* deben observar la fuerza que posee el punto de vista de que *verse rojo* es un contenido intencional al defender al realismo de los *qualia* frente al desafío funcionalista.

Una manera de caer en la falacia de la intencionalización de los *qualia* entre los realistas de los *qualia* consiste en definir los términos de las cualidades secundarias en términos de los contenidos cualitativos de las experiencias que normalmente producen —ésta es una posición compartida (pese a muchas

[9] Algunos escépticos de los *qualia* sostienen que no existe nada que sea el contenido cualitativo, mientras que otros creen que sí existe, pero que es idéntico al contenido intencional. Me parece que Dan Dennett ha vacilado sobre esta cuestión, en Dennett 1992 [incluido en esta antología, pp. 213-262], asume un punto de vista funcionalista sobre los *qualia*, al igual que en Dennett 1978b. Pero en otros escritos ha tendido a ser más eliminativista acerca de los *qualia*. Véase, en particular, Dennett 1979; y 1978c. La respuesta a la pregunta del título de Dennett 1978c (¿por qué no puede lograrse que una computadora sienta dolor?) es que no hay tal cosa como el dolor.

diferencias de opinión) por McGinn (1983) y Peacocke (1983). Peacocke define "rojo" en términos de rojo′, siendo "rojo′" el nombre que le da a "el" *quale* producido por las cosas rojas. Puesto que el contenido semántico de "se ve rojo" está dado por los correspondientes a "se ve" y a "rojo", el significado de "se ve rojo" está dado en términos de "rojo′". Esta concepción no sería muy atractiva si no existe "el" *quale* producido por las cosas rojas.

McGinn considera la posibilidad de que las cosas a las cuales tanto nosotros como los marcianos aplicamos el término "rojo" se vean para nosotros de la misma manera en que las cosas a las cuales, según ellos y nosotros, se aplica el término "verde" se ven para ellos. Aunque sigue siendo agnóstico con respecto a si dichos marcianos podrían ser totalmente idénticos a nosotros en el nivel funcional (y por lo tanto sigue siendo agnóstico con respecto a la posibilidad de un espectro invertido), concluye que, en un caso así, las cosas que son rojas relativas a nosotros (por ejemplo, los tomates maduros) son verdes relativos ellos, aunque ambos grupos apliquen "rojo" a las cosas que son del color de los tomates maduros. Luego, según McGinn, los marcianos podrían haber hablado siempre un idioma en el cual "rojo" se aplicara a las cosas del mismo modo en que nosotros lo aplicamos, pero no obstante ellos significarían verde con ella. Pero, si estamos dispuestos a considerar tal posibilidad, ¿por qué no considerar la posibilidad de que aquellos de nosotros con los lóbulos de las orejas cortos tenga una relación con aquellos con los lóbulos de las orejas largos igual a la que hay entre los terrícolas y los marcianos de McGinn? Esto no es tan absurdo como podría parecer, dado que las personas normales presentan enormes diferencias en la maquinaria fisiológica de la visión de colores (véase Hardin 1988). ¿Por qué hemos de tratar de definir "rojo" en términos de un *quale* que, supuestamente, producen las cosas rojas en todos nosotros cuando no tenemos ninguna buena razón para creer que hay algo semejante?

Mi argumento no se basa en la idea de que la inversión del espectro sea meramente posible (aunque no se dé de hecho). Sostengo que, sencillamente, no sabemos si la inversión del espectro se da o no. Aún más, incluso si encontráramos pruebas

sólidas de que la inversión del espectro de hecho no se produce, podría haber una raza de personas muy similares a nosotros, que vieran los colores, tuvieran sensaciones de colores y términos de colores, pero para los cuales sí se diera la inversión del espectro. Cabe suponer que cualquier teoría adecuada de la semántica de los términos de colores se aplicará a ellos tanto como a nosotros.[10]

Un defensor del análisis del color de McGinn y Peacocke en términos de los *qualia* podría desear admitir la posibilidad de la inversión del espectro, pero podría insistir en que, pese a todo, nuestros conceptos de colores excluyen dicha inversión. El razonamiento podría desarrollarse como sigue: antes de la teoría de la relatividad general, el concepto de una línea recta presuponía los axiomas de la geometría euclidiana, junto con la idea de que una línea recta es la distancia más corta entre dos puntos, la trayectoria que sigue una cuerda extendida o un rayo de luz, la trayectoria por la cual se utiliza menos combustible en un viaje, y cosas por el estilo. Pero la relatividad general y las observaciones pertinentes revelaron que las líneas rectas euclidianas no eran las trayectorias que poseían las otras propiedades que acabamos de listar –esas propiedades acompañaban a las líneas rectas de Riemann–, y de aquí la concepción generalmente aceptada de que el espacio es riemanniano. El meollo de la historia es que, antes del desarrollo de la relatividad general y las geometrías alternativas, separar la noción de la línea recta euclidiana del conjunto de las demás propiedades quedaba excluido de nuestros conceptos, era literalmente inconcebible, aunque no de hecho imposible. Concebirlo requirió una revolución conceptual.

La aplicación de este modelo a la inversión del espectro es ésta: la aplicación de "rojo" a las cosas rojas y la producción, por parte de las cosas rojas, del mismo *quale* en todos nosotros están unidas en nuestros conceptos, y su separación requiere una revisión conceptual. La analogía no es convincente. Al preguntarme si ves la puesta de sol tal como la veo yo, puedo imaginarte a ti teniendo una experiencia cualitativamente

[10] El argumento de Shoemaker (1982) contra Harrison propicia esta observación.

igual a la mía. En la vida cotidiana, usualmente no tomamos en cuenta la posibilidad de una inversión del espectro. Pero sería un error saltar de este hecho acerca de nuestras prácticas a la idea de que la inversión del espectro está excluida por nuestros conceptos. Los niños pueden entender fácilmente la inversión del espectro (suponiendo que pueda ser entendida). Inténtalo con el próximo chico de ocho años que se cruce en tu camino. En realidad, a veces los niños llegan a esta idea por sí solos. (Mi hija de once años propuso la inversión del espectro para explicar por qué la gente difiere en sus colores predilectos.)

Mi versión del realismo de los *qualia* es cuasifuncional. En mi opinión, el contenido **intencional** de la experiencia es funcional. Una experiencia posee el contenido intencional de verse rojo si funciona correctamente, si es causada por cosas rojas en las circunstancias apropiadas y usada adecuadamente al pensar acerca de las cosas rojas y al actuar con respecto a las cosas rojas.[11] Los papeles funcionales de los que estoy hablando son los que denomino papeles "de largo alcance", papeles que incluyen a las cosas reales del mundo como entradas y salidas. Deben ser distinguidos de los papeles "de corto alcance" que en ocasiones prefieren los funcionalistas, papeles que se detienen en la piel. Para el papel funcional que caracteriza el contenido intencional de *verse rojo* es esencial que sea causado (apropiadamente) por las cosas rojas y que cause pensamientos apropiados de las cosas rojas, y acciones apropiadas a ellas.

Tal es la razón de que mi versión del realismo de los *qualia* sea cuasi**funcional**; y he aquí por qué es **cuasi**funcional: el contenido **cualitativo** de la experiencia **no** puede caracterizarse en términos funcionales. Dos experiencias pueden diferir funcionalmente, y por ende tener contenidos intencionales distintos, pero pueden poseer el mismo contenido cualitativo, es decir, ser iguales con respecto a "cómo es" experimentarlas. Aún más, dos experiencias pueden ser iguales en cuanto a su función (y

[11] Nótese que he dicho que la función de **una experiencia** es lo que le confiere el contenido intencional de *verse rojo*; no he dicho que cualquier cosa (incluso algo que no sea una experiencia) que posea esa función tiene ese contenido. Esta cuestión –la de los "*qualia* ausentes"– no se revisa en este artículo.

por ende tener el mismo contenido intencional), pero tener contenidos cualitativos diferentes.

De este modo, el realismo cuasifuncional de los *qualia* es funcionalista acerca del contenido intencional de la experiencia. Y también es funcionalista acerca de la creencia de que la sangre es roja, del concepto de rojo y del significado de "rojo".[12]

Pero, ¿no identifiqué antes los "conceptos" del color con los contenidos cualitativos? Sí, y ésta constituye indudablemente una manera razonable de utilizar el concepto de un concepto, pero uno puede individuar igualmente bien los conceptos de acuerdo con los significados del lenguaje público, como lo hace Harman. Él dice que lo que hace que un concepto sea el concepto de rojo consiste en que es producido de un modo natural e inmediato por cosas rojas, en el papel que desempeña en el pensamiento y la manipulación de cosas rojas, en que causa que el sujeto juzgue que dos cosas son del mismo color, y demás. A este respecto, el realista cuasifuncional de los *qualia* bien puede estar de acuerdo con Harman. Así, tomemos como base común que la posesión del concepto de rojo es una cuestión de cómo funcionan nuestras experiencias, y no de sus propiedades cualitativas intrínsecas.[13]

La refutación del espectro invertido planteada por Harman depende de que se rechace la distinción de la que he estado hablando entre el contenido cualitativo y el intencional. Afirma que la experiencia posee sólo un tipo de contenido: el contenido intencional. Según Harman, dos experiencias con el mismo contenido intencional deben ser iguales en todos sus aspectos mentales. Su rechazo de la distinción lo lleva a formular y refutar una versión de la hipótesis del espectro invertido que incluye únicamente el contenido intencional. Su versión de la hipótesis del espectro invertido es como sigue: "las cosas que le parecen

[12] Pero por las razones mencionadas previamente, no es funcionalista acerca del significado de "cómo es" y de "contenido cualitativo". La posición que estoy defendiendo sobre la semántica de los términos de colores se parece a la de Paul Churchland, en Churchland 1979, cap. 2.

[13] Siempre que la experiencia posea **algún** contenido cualitativo que, como experiencia, tiene que poseer. Como señalé en la nota anterior, no pretendo discutir aquí si algo que no es una experiencia puede tener el papel funcional de una experiencia.

rojas a una persona le parecen verdes a la otra, las cosas que le parecen anaranjadas a la primera persona le parecen azules a la segunda, y así sucesivamente", donde las personas a las que se refiere son iguales desde el punto de vista funcional.[14] Pero esta hipótesis del espectro invertido es una construcción de paja. De acuerdo con esta hipótesis de paja del espectro invertido, la experiencia que obtienes al ver cosas rojas, tu experiencia como de rojo, la que produce proferencias sinceras como "Esto es rojo" de manera normal, esta experiencia podría tener el contenido intencional de *verse verde*. Pero ninguno de los que proponen el espectro invertido debería aceptar **esta** hipótesis del espectro invertido. Los proponentes del espectro invertido postulan contenidos cualitativos invertidos, no contenidos **intencionales** invertidos.

Quizá el malentendido en este terreno sea parcialmente responsable del hecho de que ambos bandos del argumento sobre el espectro invertido tiendan a no ver nada en la parte contraria. Por citar un escrito anterior de Harman, donde trata el tema:

> Hablo aquí de un "argumento", aunque (como lo ha observado D. Lewis en un contexto similar) el "argumento" se reduce simplemente a negar la explicación funcionalista del contenido de conceptos y pensamientos, sin ofrecer realmente ninguna razón de esa negación.[15]

La hipótesis de paja del espectro invertido tiene efectivamente este sabor a petición de principio, suponiendo, como lo hace, que pudiera haber contenidos intencionales invertidos. Suponer que pudiera haber contenidos intencionales invertidos equivale a suponer que el concepto de rojo que tiene una persona podría funcionar como el concepto de verde, sin por ello ser un concepto de verde. Y esto niega precisamente la explicación funcionalista de los conceptos.

[14] Lycan 1987 también presenta la hipótesis del espectro invertido de esta manera.

[15] Harman 1982, p. 250. Lycan 1973, p. 60, dice que los proponentes de la hipótesis del espectro invertido "simplemente [...] niegan la verdad del funcionalismo... sin ningún argumento".

2. *El espectro invertido y la Tierra Invertida*

Consideremos algunos argumentos positivos con respecto a la posibilidad de un espectro invertido. Si el proponente del espectro invertido pudiera establecer que el peso de la prueba recae en quien niega su posibilidad, entonces los argumentos positivos serían innecesarios. Pero, dado que nadie ha logrado establecer el peso de la prueba, en la literatura se enfatiza lo **epistémico**: se producen y se discuten casos de ciencia ficción que (supuestamente) contarían como **evidencia** en favor de un espectro invertido.[16] La idea consiste en que si pudiera haber evidencia en favor de un espectro invertido, éste sería entonces posible. Por ejemplo, imaginemos a un par de gemelos genéticamente idénticos, a uno de los cuales, al nacer, se le implantaron en los ojos lentes que invierten los colores.[17]

[16] Véase, por ejemplo, Shoemaker 1982 y Lycan 1987.

[17] Quizá una posibilidad más realista sería la de una computadora miniatura implantada en la cavidad del seno nasal y conectada al nervio óptico, detrás de la retina, de suerte que registre las salidas de la retina y cambie las señales que representan el rojo a señales que representen el verde, y así con los demás colores, transmitiendo las señales transformadas al núcleo geniculado lateral. Muchas de las objeciones a la posibilidad empírica de un espectro invertido planteadas por Harrison y Hardin también son objeciones a la posiblidad de un inversor de los colores. Véase Harrison 1973 y Hardin 1988. Véase también el comentario de Shoemaker sobre Harrison en el artículo citado anteriormente (Shoemaker 1982, y p. 336 de Shoemaker 1984). Tanto Harrison como Hardin argumentan de distinta manera que el color sólido posee asimetrías que no permiten que la inversión de colores pase sea indetectable. Como se pondrá de manifiesto más adelante, la afirmación principal de este artículo no resultaría perjudicada si ellos tuvieran razón, pues sólo estoy alegando que el argumento de la "Tierra Invertida" que presentaré es mejor argumento contra el funcionalismo que el argumento del espectro invertido. (Además, en ambos argumentos puede evitarse recurrir a anteojos inversores.) Sin embargo, no puedo dejar de mencionar que ni las objeciones de Harrison ni las de Hardin sirven para el caso. Hardin señala los problemas principales con Harrison. Por ejemplo, Harrison argumenta que puede haber más tonos entre ciertos colores (primarios) particulares que entre otros, lo cual provoca que la inversión sea detectable. Hardin señala que la gama de matices que se observan entre dos colores cualesquiera varía mucho entre personas que ven los colores de manera perfectamente normal. La objeción de Hardin es que algunos colores (el rojo y el amarillo) son cálidos, mientras que otros (el azul y el verde) son fríos. Pero no ofrece ningún argumento contra la posibilidad de una inversión entre los colores cálidos y fríos. Esto es, no presenta un ar-

Los dos gemelos son criados normalmente y, al llegar a la edad adulta, ambos aplican la palabra "rojo" a las cosas rojas de la manera normal. Pero aunque los gemelos son funcionalmente idénticos en los aspectos pertinentes, podemos suponer que el estado fisiológico interno que media entre las cosas rojas y las proferencias de "rojo" en uno de ellos es igual al estado fisiológico interno que media entre las cosas verdes y las proferencias de "verde" en el otro. Y a partir de esto uno podría argumentar en favor de la afirmación de que las cosas que ambos llaman rojas le parecen a uno de la misma manera en que las cosas que ambos llaman verdes le parecen al otro. Hay mucho que decir por ambas partes del debate. No deseo entrar aquí en la discusión.

Los casos interpersonales de espectro invertido como el que acabamos de mencionar no han sido tomados muy seriamente por la oposición. Ésta supone (como lo hace Harman) que no hay razones para pensar que diferentes realizaciones fisiológicas de la experiencia de cosas rojas implique alguna diferencia **experiencial**. Como cualquier estado mental, la experiencia de

gumento en contra de la posibilidad de que las cosas que, según coincidimos, poseen un color cálido producen el mismo efecto cualitativo en ti que las cosas que, según coincidimos, poseen un color frío producen en mí.

Hardin también arguye que si la persona que ve invertidos los colores ve la sangre como verde y fría, el epifenomenalismo tendría que ser verdadero, mientras que la suposición de que ve la sangre de color verde y cálido raya en la incoherencia. Estas afirmaciones –por lo menos la primera– derivan su plausibilidad de la falacia de la intencionalización de los *qualia*. La razón de que la posibilidad de ver la sangre como verde y fría implique, supuestamente, el epifenomenalismo estriba en que se supone que ver la sangre como de color frío llevaría a decir y hacer cosas como si fuera de color frío, lo cual resulta incompatible con la identidad funcional (y por lo tanto conductual) presupuesta en la hipótesis del espectro invertido. Pero esta afirmación depende de que se considere al espectro invertido como una inversión intencional. Supongamos que Q_a es la propiedad cualitativa que te lleva a decir que algo es rojo y que me lleva a mí a decir que algo es verde, mientras que Q_b posee el efecto opuesto. Y supongamos que Q_x es la propiedad cualitativa que te lleva a decir que algo es de color cálido y que me lleva a mí a decir que algo es de color frío, mientras que Q_y posee el efecto opuesto. No hay amenaza de epifenomenalismo en que yo experimente las cosas rojas a través del contenido cualitativo $Q_b + Q_y$. La apariencia de un problema de epifenomenalismo sólo proviene de la falacia de la intencionalización de los *qualia*, al pensar que $Q_b + Q_y$ posee un contenido intencional que conlleva frialdad.

rojo tiene realizaciones fisiológicas alternativas, y se afirma que éste es sólo un caso de realizaciones alternativas de exactamente la misma experiencia.

La respuesta que el realista de los *qualia* ofrece a Harman es que cuando **nosotros** nos ponemos lentes que invierten los colores, la hierba se ve roja, las fresas maduras verdes, el cielo amarillo, y así sucesivamente. ¿No hemos de suponer entonces que lo mismo se aplica al gemelo que lleva los lentes inversores? En este punto, como en otros en las discusiones acerca de los *qualia*, haremos bien en seguir el consejo de Searle (1992) de fijarnos en el punto de vista de la primera persona. Recurrir a lo que sucede cuando nosotros nos colocamos los lentes inversores apunta a una versión del ejemplo del espectro invertido que entraña un caso semejante: el espectro invertido **intra**personal.[18]

Pienso que es mejor tratar el espectro invertido intrapersonal como si constara de cuatro etapas. Primera: tenemos a una persona funcionalmente normal. Segunda: se le implantan lentes inversores en los ojos y el sujeto dice que la hierba se ve roja y la sangre verde.[19] Tercera: después de un periodo de usos confusos de los términos de colores, se adapta finalmente al grado de usar normalmente el lenguaje de los colores. Esto es, describe de manera natural e inmediata la sangre como "roja" y la hierba como "verde". En esta tercera etapa, es funcionalmente normal salvo en un aspecto importante: recuerda el tiempo previo a la implantación de los lentes como la época en la que "la hierba se veía igual a como hoy veo la sangre". Cuarta: el sujeto sufre amnesia con respecto al periodo anterior a la implantación de los lentes y se vuelve una persona totalmente normal desde el punto de vista funcional, tal como en el primer periodo.

[18] Dije que cuando nos colocamos los lentes inversores de los colores, la hierba se ve roja; así que, ¿no estaré comprometiéndome con la afirmación de que el gemelo que usa lentes inversores también ve roja la hierba? No. Para el usuario **temporal** de los lentes la hierba se ve igual que como normalmente se ven las cosas rojas. Pero el gemelo invertido no es un usuario temporal de los lentes: para él la hierba se ve igual que como normalmente se ven las cosas verdes.

[19] En Shoemaker 1982, esta etapa está dividida en partes a fin de aumentar su poder de convencimiento.

La etapa crucial es la tercera. En ella tenemos la evidencia del testimonio de la víctima de que la hierba solía verse tal como ahora se ve la sangre. Obsérvese que, según he descrito el caso, no aparece la palabra "*quale*" ni otros términos técnicos en lo que dice el sujeto; se limita a hablar de cómo se ven las cosas rojas y cómo solían verse en comparación con las cosas verdes. En esta etapa, es una persona funcionalmente anormal precisamente porque hace esas extrañas afirmaciones. Pero si estamos dispuestos a creerle, entonces cuando sufre amnesia y se vuelve funcionalmente normal, ¿por qué habríamos de pensar que sus *qualia* se han vuelto a invertir?

La ventaja principal del caso intrapersonal frente al interpersonal estriba en que se dispone del informe introspectivo del sujeto durante la tercera etapa. Supongamos que **tú** eres el sujeto. De acuerdo con la hipótesis, recuerdas cómo se veía antes la hierba, y te das cuenta de que solías verlo igual a como ahora se ven los tomates maduros. ¿Hay en esto algo incoherente? Dennett y Rey dicen que sí. Dennett insiste en que tu inclinación por decir esas cosas refleja un malfuncionamiento de la memoria, y Rey señala que no sabemos cómo entender lo que dices.[20] Admito que estas objeciones son justas en cierta medida (aunque no estoy de acuerdo con ellas). Después de todo, el sujeto adaptado no es una persona normal. Quizá pueda probarse que la hipótesis de la confusión siempre será más plausible que la hipótesis de la inversión del espectro. Indudablemente, si se piensa que la distinción entre contenido intencional y cualitativo constituye una confusión, lo natural será buscar alguna forma de evitar aceptar sin más los reportes de los recuerdos del sujeto cuando está en la tercera etapa.

Creo que estas críticas pueden refutarse con sus propios términos, pero en vez de intentar hacerlo aquí, me propongo tomar un rumbo distinto. Describiré un caso que es el "opuesto" al caso habitual del espectro invertido, un caso de contenido **intencional** invertido y de inversión funcional combinados con un contenido cualitativo idéntico. En el caso típico del espectro invertido, tenemos a dos personas (o etapas de la misma persona)

[20] Véase Dennett 1992 [incluido en esta antología, pp. 213–262], y Rey (inédito).

cuyas experiencias son iguales funcional e intencionalmente, pero que están invertidas en el nivel cualitativo. Describiré un caso de dos personas o etapas cuyas experiencias son cualitativamente iguales, pero intencional y funcionalmente invertidas. Si estoy en lo correcto acerca de este caso, la distinción entre el contenido intencional y el cualitativo de la experiencia quedará reivindicada, y se refutará la teoría funcionalista del contenido cualitativo.

Tal como sucede con el argumento habitual del espectro invertido, el mío depende de un ejemplo de ciencia ficción. (La fantasía en mi ejemplo podría reducirse en gran medida. Por ejemplo, imaginemos, en vez de planetas enteros, habitaciones en las que nuestros sujetos están encerrados de por vida.) Recurriré a un ejemplo de Harman: la Tierra Invertida.[21] La Tierra Invertida se diferencia de la Tierra en dos aspectos. En primer lugar, todo tiene el color complementario al color que posee en la Tierra. El cielo es amarillo, la hierba es roja, las tomas de agua contra incendios son verdes, etc. Quiero decir que todas las cosas tienen **realmente** estos excéntricos colores. Si visitaras la Tierra Invertida junto con un equipo de científicos de tu universidad, todos estarían de acuerdo en que en este planeta el cielo es amarillo, la hierba es roja, etc. En segundo lugar, el vocabulario de los habitantes de la Tierra Invertida también está invertido: si les preguntan de qué color es el cielo (amarillo), dirán (verazmente) "¡Azul!" Si les preguntan de qué color es la hierba (roja), dirán "Verde". Si llevaran a una persona que habla el dialecto de la Tierra Invertida a un sitio neutral (donde el cielo tiene un color desconocido, la vegetación es extraña, y demás) y contrataran a un equipo de lingüistas para que, echando mano de todos los métodos razonables para examinar su idioma, tendrían que llegar a la conclusión de que esa persona utiliza "rojo" para significar lo que nosotros significamos con "verde", "azul" para significar lo que nosotros significamos con "amarillo", etc. Tendrían que llegar a la conclusión de que el dialecto de la Tierra Invertida se distingue

[21] Harman 1982, p. 251. Tengo un ejemplo anterior, aunque más torpe, a este respecto en Block 1980a, pp. 302–303, n. 21. Véase también el uso que Ron McClamrock da al ejemplo de Harman para un propósito distinto en su tesis de doctorado de 1984.

del nuestro en que las palabras de colores tienen "significados invertidos". Si se desarrollara el comercio entre los dos planetas y los pintores de la Tierra Invertida pidieran pintura a una de nuestras tiendas, tendríamos que traducir su pedido de "pintura verde" en un pedido al empleado del almacén para que les diera pintura roja. La Tierra Invertida se diferencia de la Tierra por la permuta de términos y la permuta de estímulos.

Aún más, los contenidos intencionales de las actitudes y de las experiencias de los habitantes de la Tierra Invertida también se encuentran invertidos. Si un extranjero lee equivocadamente un diccionario fraseológico y llega a la conclusión de que "bote de basura" significa un tipo de sándwich, a pesar de lo que diga, en realidad no quiere almorzar un bote de basura —quiere un sándwich. Asimismo, cuando un habitante de la Tierra Invertida se pregunta, en sus propios términos, "¿Por qué el cielo es azul?", no se está preguntando por qué el cielo es azul, sino por qué es amarillo. Y cuando mira el cielo (en circunstancias normales para la Tierra Invertida), la experiencia que tiene y que describiría como el cielo se ve "azul" es de hecho la experiencia del cielo viéndose amarillo. No hay aquí misterio ni asimetría. Estoy utilizando nuestro lenguaje para describir sus contenidos intencionales. Si él empleara su lenguaje para describir los nuestros, nos describiría correctamente como invertidos de la misma forma.

Estas dos diferencias que he mencionado "se neutralizan" en el sentido de que la manera de hablar de la Tierra Invertida **sonará** exactamente igual a la manera de hablar terrestre. Los programas de radio de la Tierra Invertida sonarán como los de lugares lejanos donde se habla inglés, como Nueva Zelandia. Los niños de ambos planetas les preguntarán a sus padres en su idioma: "¿Por qué es azul el cielo?" "¿Por qué la hierba es verde?"

Presentemos ahora el caso análogo al ejemplo del espectro invertido intrasubjetivo. Un equipo de científicos locos te dejan inconsciente. Mientras estás inconsciente, te implantan en los ojos lentes inversores de colores y te modifican los pigmentos corporales para que no sufras una sorpresa desagradable al despertar y mirarte los pies. Te transportan a la Tierra Invertida, donde te intercambian por tu contrapartida que ha ocupado

en la Tierra Invertida un lugar que corresponde exactamente (salvo por el color de las cosas) a tu lugar en tu tierra de origen. Te despiertas, dado que los lentes neutralizan la inversión de colores, no notas ninguna diferencia. El "cómo es" para ti interactuar con el mundo y con las demás personas no sufre tampoco cambio alguno. Por ejemplo, ves azul el cielo amarillo, y toda la gente que te rodea describe los objetos amarillos, tales como el cielo, como "azules". En lo que concierne al aspecto cualitativo de tu vida mental, no hay nada distinto de la forma en que habrían sido las cosas de haber permanecido en tu tierra natal. Incluso podemos suponer que tu cerebro es exactamente igual, en cuanto a sus propiedades fisiológicas, como habría sido si te hubieras quedado en tu lugar de origen. (Desde luego, los enunciados de la ciencia del color tendrían que ser distintos en ambos planetas. Nosotros describimos como "amarilla" la luz pura de 577 nanómetros, mientras que ellos la describirían como "azul". Así, supongamos que tú y los gemelos de los ejemplos que siguen no saben nada de la ciencia del color y, por ende, que estas diferencias científicas no les afectan.)[22]

Ahí estás en la Tierra Invertida. El contenido cualitativo de tu cielo (local) es exactamente igual al del día previo, cuando mirabas el cielo en tu tierra natal. ¿Qué hay del contenido intencional de esta experiencia? A este respecto pueden suscitarse algunos desacuerdos. Las bases causales de tus palabras de colores se adecúa casi por completo al de la tierra de origen; tu uso del término "azul" está fundado en las referencias a las cosas azules y a los usos que otras personas hacen de "azul" para referirse a las cosas azules. Por esta razón, yo diría que, durante tu primer día en la Tierra Invertida, tus contenidos intencionales siguen siendo los mismos que antes, es decir, diferentes de los nativos. Al principio, cuando mires el cielo pensando el pensamiento que expresarías como "Es tan azul como siempre", estarías expresando el mismo pensamiento que habrías expresado ayer en la Tierra, sólo que hoy estás **equivocado**.

[22] En mi ejemplo más torpe (mencionado en la nota previa), recurría a los lentes inversores y a la permuta de términos de colores, pero en lugar del cielo amarillo y demás objetos de la Tierra Invertida, hice que los sucesos se dieran en una apartada aldea ártica donde (supuse) la gente carecía de creencias constantes acerca de los colores de las cosas.

Además, tu pensamiento no es igual al que un nativo de la Tierra Invertida expresaría con las mismas palabras. No obstante, en mi opinión, después de pasar suficiente tiempo en la Tierra Invertida, tu compenetración con el nuevo entorno físico y lingüístico dominaría, de modo que tus contenidos intencionales variarían hasta llegar a ser iguales a los contenidos intencionales de los lugareños. Considérese una analogía (ofrecida por Martin Davies): si tuvieras la capacidad de reconocer a Margaret Thatcher [MT] antes de tu viaje a la Tierra Invertida y al llegar identificas erróneamente a la gemela MT' como MT, estás equivocado. Sin embargo, con el tiempo, tus juicios como "Ella es MT" terminan por ser acerca de la gemela MT y se convierten en juicios correctos tras haber comenzado como erróneos. Si fueses raptado a la edad de quince años, al cabo de cincuenta años utilizarías "rojo" para significar verde, tal como hacen los nativos. Una vez invertidos tus contenidos intencionales, también se invierten tus estados funcionales. El estado que ahora causan **normalmente** las cosas azules es el mismo estado que antes causaban normalmente las cosas amarillas. De manera que después de transcurridos cincuenta años, tú y tu etapa anterior en la Tierra ejemplificarían lo que yo pretendo, a saber, un caso de inversión funcional e intencional junto con los mismos contenidos cualitativos —lo opuesto al caso del espectro invertido. Esto basta para refutar la teoría funcionalista del contenido cualitativo y, al mismo tiempo, para establecer la distinción entre lo intencional y lo cualitativo.

Pero, ¿qué ocurriría si los hechos de la fisiología humana se interpusieran en el desarrollo del caso según lo he descrito? Mi respuesta es la misma que la mencionada anteriormente (fundada en los argumentos con los que Shoemaker rebate a Harrison), a saber, que es posible que exista una raza de personas muy semejante a nosotros, con visión de colores y sensaciones de colores, pero cuya fisiología no descarte el caso expuesto (o la inversión del espectro). El funcionalista difícilmente se sentiría satisfecho con la afirmación de que nuestras experiencias son estados funcionales, pero que las experiencias de la otra raza no lo son.

La teoría funcionalista que tengo en mente es una teoría de la identidad de estados funcionales, una teoría que dice que

todos y cada uno de los estados mentales **es** algún estado funcional. Si el estado mental M = el estado funcional F, entonces cualquier instancia de M debe ser también una instancia de F. Pero, en el caso de la Tierra Invertida que acabo de describir, dos instancias del mismo estado cualitativo —el tuyo y el de tu Gemelo— desempeñan papeles funcionales diferentes. Y el estado cualitativo que compartes con tu Gemelo no puede ser idéntico **tanto** a tu estado funcional **como** al estado funcional del Gemelo, ya que éste sería un caso de $Q = F_1$, $Q = F_2$ y $\sim (F_1 = F_2)$, lo cual contraviene la transitividad de la identidad. Este argumento tiene aún un cabo suelto: la existencia de un tipo de descripción funcional según la cual los dos estados cualitativos son funcionalmente distintos no descarta la posibilidad de otro tipo de descripción funcional según la cual sean iguales. Volveré a esta cuestión más adelante, cuando examine las objeciones al argumento sobre la Tierra Invertida.

Obsérvese que este caso intrapersonal de la Tierra Invertida no tiene el punto débil que señalamos en el caso intrapersonal del espectro invertido. En este último, la perturbación interna del sujeto hace que sus reportes en primera persona estén expuestos a dudas. Pero tú, el sujeto del caso de la Tierra Invertida, no has tenido ninguna perturbación interna. Tu traslado a la Tierra Invertida se realizó sin que lo notaras, —no hubo un periodo de confusión o adaptación. Toda la inversión se produjo fuera de tu cerebro: en tus lentes inversores y en tu entorno físico y lingüístico. El caso intrapersonal posee la considerable ventaja de servir como testimonio tanto de la Tierra Invertida como del caso del espectro invertido. Pero algunos lectores podrían abrigar ciertas dudas con respecto a la afirmación de que al cabo de cincuenta años tus contenidos intencionales y tu lenguaje sean iguales a los de los nativos de la Tierra Invertida. Por esta razón, quizá resulte útil pasar brevemente al caso interpersonal de la Tierra Invertida. Hablaré en términos del caso interpersonal más adelante, pero espero que el lector se remita de vez en cuando al punto de vista de la primera persona que está mejor captado en el caso intrapersonal.

Imaginemos a un par de gemelos genéticamente idénticos, uno de los cuales es adoptado en la Tierra Invertida después de que se le han implantado lentes inversores y modificado los pig-

mentos corporales; y el otro recibe su crianza normalmente en la Tierra (sin lentes especiales). A fin de introducir el punto de vista de la primera persona, imaginemos que tú eres el gemelo que se quedó en la Tierra. Cuando ustedes dos miran sus respectivos cielos, ambos oyen "azul" y reciben el mismo estímulo retinal. Podemos suponer que tú y el Gemelo son iguales en los aspectos neurológicos pertinentes. (Incluso podríamos suponer que sus cerebros son duplicados moleculares.) Pero entonces ambos serían iguales con respecto a los contenidos cualitativos de sus experiencias, igual que en el ejemplo intrasubjetivo, en el que tus contenidos cualitativos eran los mismos antes y después del viaje a la Tierra Invertida; pues los contenidos cualitativos sobrevienen en la constitución física, y la constitución física de ambos gemelos es esencialmente igual.

Aunque tú y tu Gemelo tienen los mismos contenidos cualitativos en todo momento, están invertidos en el nivel funcional y con respecto a los contenidos intencionales de la experiencia; pues con respecto a sus contenidos intencionales, el Gemelo es igual a sus amigos de la Tierra Invertida. Cuando el Gemelo o cualquier otro habitante de la Tierra Invertida tiene el pensamiento que expresaría como "El azul del cielo está muy saturado", su pensamiento es acerca del **amarillo**. Luego, los contenidos intencionales de tus experiencias están invertidos con respecto a los del Gemelo. ¿Qué ocurre entonces con la inversión funcional?

La representación mental que uno tiene de azul —el concepto que uno tiene de azul, en el sentido que los psicólogos dan al término "concepto"— desempeña un papel conocido al ser producido de manera natural e inmediata por cosas azules y al controlar el comportamiento que uno tiene hacia las cosas azules. (Harman supone que la experiencia que uno tiene del azul **es** el concepto de azul que uno tiene; no me molesta seguirlo en esto; la cuestión no es pertinente en la controversia aquí discutida.) Estoy hablando del conocido papel del concepto que uno tiene de azul. Quizá en este punto resultaría útil pensar que tú y el Gemelo están jugando un juego. Se les muestra algo de cierto color y su tarea consiste en encontrar otra cosa del mismo color. Te muestro una fresa madura, y miras a tu alrededor, y encuentras un autobús inglés de dos pisos con el que

hace juego. El Gemelo, en su situación correspondiente, aparea sus fresas verdes con un autobús verde.

Ahí están tú y el Gemelo, contemplando sus respectivos cielos, teniendo experiencias con exactamente los mismos contenidos cualitativos. En este preciso momento, te encuentras en un estado producido de manera natural e inmediata por cosas azules, un estado que desempeña el ya mencionado y conocido papel funcional en el control de tus respuestas a las cosas azules. En el mismo momento, el Gemelo se encuentra en un estado producido por cosas amarillas, un estado que desempeña el mismo y conocido papel funcional en el control de sus respuestas a las cosas amarillas, de suerte que estás funcionalmente invertido con respecto al Gemelo. Piensa en el juego de aparear colores. El estado cualitativo que funciona para guiarte en tu tarea de aparear una cosa azul con una muestra azul también funciona para guiar al Gemelo en su tarea de aparear una cosa amarilla con una muestra amarilla. Desde luego, existe una considerable similitud funcional entre el Gemelo y tú; de hecho, ambos son exactamente iguales en el nivel funcional si consideramos que la frontera entre lo interno y lo externo está en la piel, incluidas las emisiones del mismo sonido, "¡Otra vez azul!", cuando ven el cielo. Más adelante retomaré la cuestión de si esta descripción funcional interna puede utilizarse para evitar la crítica al funcionalismo que estoy desarrollando.

Aún más, tu estado perceptual es acerca de las cosas azules, mientras que el del Gemelo es acerca de las cosas amarillas, de modo que se encuentran intencionalmente invertidos. (La "inversión" está relacionada con el hecho de que el amarillo y el azul son simplemente ejemplos de colores complementarios.)

Obsérvese que no estoy diciendo que si tú y el Gemelo se presentaran juntos en mi estudio, estarían entonces intencional o funcionalmente invertidos. (Al mostrárseles una cosa roja y preguntárseles su color, ambos emitirían sonidos diferentes, pero significarían lo mismo con ellos.) El caso de la Tierra Invertida **depende** de que ustedes no estén juntos en mi estudio. La inversión ambiental es crucial para este caso, de manera que no sirve de nada imaginar que ha desaparecido. A este respecto, el caso de la Tierra Invertida resulta un poco menos firme para la

imaginación que el caso del espectro invertido. Se puede imaginar a **esos** gemelos en **cualquier** entorno, conservando aun la misma identidad funcional y la misma inversión cualitativa. Pero esto no constituye una desventaja argumental real para el caso de la Tierra Invertida. El lector sólo debe recordar que la inversión ambiental forma parte de lo que hace que este ejemplo funcione, y por eso es preciso conservarla.

El resultado es que si tú y el Gemelo contemplan sus respectivos cielos, diciendo "Todavía es azul", están teniendo experiencias que poseen el mismo contenido cualitativo, pero un contenido intencional invertido, y que se encuentran funcionalmente invertidas. En conclusión, la distinción entre contenido intencional y cualitativo se reivindica, y se refuta la teoría funcional de los *qualia*.

Es posible atacar de diversas maneras el argumento de la Tierra Invertida que acabo de exponer. Examinaré brevemente algunas de ellas, pero antes quiero hacer ciertas comparaciones con el argumento más tradicional del espectro invertido.

Se pretende que el caso de la Tierra Invertida ejemplifique la identidad cualitativa junto con la inversión funcional e intencional, mientras que se pretende que el caso del espectro invertido ejemplifique lo contrario: la inversión cualitativa junto con la identidad funcional e intencional. Como ya lo mencioné, la diferencia cualitativa en la mejor versión del caso del espectro invertido (la inversión intrasubjetiva) intenta ser establecida por medio del testimonio del sujeto que sufrió la inversión. Como también lo señalé, el punto débil del caso estriba en que han sucedido cosas inusuales en el interior de este sujeto (el periodo de "adaptación"), y por lo cual se suscitan dudas (que algunos filósofos toman más seriamente que otros) acerca de si las palabras del sujeto pueden ser creídas sin más. El contraste con el caso de la Tierra Invertida es notable. Cuando eres raptado y despiertas en la Tierra Invertida, no existe diferencia alguna en cómo son las cosas para ti. En este caso, al contrario del espectro invertido intrasubjetivo, no se produce una perturbación interna que arroje dudas sobre el testimonio del sujeto. (Desde luego, todo sujeto puede extrañar una serie de cambios menores, pero no hay motivos para sospechar de ellos en el caso de la Tierra Invertida.) Además, en el caso

intersubjetivo de la Tierra Invertida, tenemos a la superveniencia de nuestro lado. Tu cerebro es un duplicado molecular del cerebro del Gemelo, y dado que las propiedades intrínsecas como el contenido cualitativo supuestamente sobrevienen en las propiedades físicas, podemos concluir que eres cualitativamente idéntico al Gemelo. Observemos que la superveniencia es asimétrica con respecto a la igualdad y la diferencia. Esto es, la identidad molécula por molécula entre su cerebro y el del Gemelo establece que propiedades intrínsecas como los *qualia* sean iguales. Pero la diferencia física entre los cerebros de los gemelos del espectro invertido no establece ninguna diferencia cualitativa entre ellos, porque el crítico siempre puede aducir (como lo hace Harman) que la diferencia física no es más que una cuestión de realizaciones alternativas del mismo estado mental.

La ventaja del caso de la Tierra Invertida consiste en que no hay una permutación **interna** como en el caso del espectro invertido. La permuta tiene lugar fuera del cerebro. Hasta aquí el argumento de la Tierra Invertida resulta un mejor argumento contra el funcionalismo que el argumento del espectro invertido. Pero, ¿sale tan bien librado de las objeciones?

La primera objeción que quiero considerar es que el Gemelo no tiene los mismos contenidos intencionales de sus amigos de la Tierra Invertida porque cuando mira el cielo tiene una experiencia cualitativamente distinta de la que tienen sus amigos (porque él tiene lentes inversores y ellos no). Y los juicios del Gemelo acerca de los colores responden a los contenidos cualitativos de su experiencia. Así, cuando él dice y sus amigos repiten "El cielo se ve azul", lo que ellos dicen es verdadero mientras que lo que él dice es falso, porque para **él** el cielo no se ve azul.[23]

Lo primero que debe notarse con respecto a esta objeción es que supone un espectro invertido; afirma que el espectro del Gemelo está invertido en relación con el de sus amigos, y utiliza

[23] Esta objeción está sugerida en la idea de McGinn y Peacocke mencionada anteriormente, según la cual el contenido intencional ("el contenido representacional" de acuerdo con la terminología de Peacocke) de una experiencia como de rojo se define en términos de su contenido cualitativo (contenido sensorial).

esta afirmación para argüir que los contenidos intencionales del Gemelo no son los mismos que los de sus amigos. Por esta razón, ningún funcionalista puede aceptar esta objeción, y Harman menos que nadie. Harman insiste (como yo lo he hecho) en que el contenido intencional de la experiencia como de amarillo consiste en que es producida de manera natural e inmediata por cosas amarillas, en su papel en la manipulación de cosas amarillas y demás. Y puesto que el Gemelo es igual que sus amigos en este aspecto, Harman no puede suponer que difiere de ellos en cuanto al contenido intencional.

Además, quienes aceptamos la posibilidad de un espectro invertido no podemos aceptar tampoco esta objeción, ya que incurre en la falacia de la intencionalización de los *qualia*. Si los espectros invertidos son posibles, los nativos de la Tierra Invertida pueden diferir entonces unos de otros en lo que respecta al contenido cualitativo producido por el cielo. Ninguno de ellos puede ser señalado como el único cuyo contenido cualitativo al contemplar el cielo es **realmente** el de *verse amarillo*. No se puede reducir el contenido intencional al contenido cualitativo.

Obsérvese que no he recurrido a la posibilidad de un espectro invertido para defender a la Tierra Invertida. La objeción supone un espectro invertido, y yo me he limitado a indicar las consecuencias de ese supuesto.

Pasemos ahora a una objeción más seria. He dicho que eres cualitativamente idéntico al Gemelo, pero funcionalmente invertido con respecto a él. En el momento en que ambos contemplan sus cielos respectivos, te encuentras en un estado producido por cosas azules, el cual controla la manipulación de cosas azules, mientras que el Gemelo se encuentra en un estado producido por cosas amarillas, el cual controla la manipulación de cosas amarillas. (Recordemos el ejemplo del juego de aparear colores.) Mi razonamiento conlleva pensar en las entradas y salidas como si involucrasen a los objetos del mundo. Pero esto no es más que una forma de pensar en las entradas y salidas de acuerdo con un tipo de descripción funcional. La existencia de una descripción funcional según la cual tú eres una inversión del Gemelo no excluye otras descripciones funcionales según las cuales tú y el Gemelo serían funcionalmente iguales. Si hay

otras descripciones funcionales según las cuales tú y el Gemelo son funcionalmente iguales, ¿cómo puedo afirmar que he refutado el funcionalismo?

De hecho, hay una variedad de descripciones funcionales según las cuales eres funcionalmente igual al Gemelo. Por ejemplo, tu cerebro es físicamente idéntico al del Gemelo, de manera que cualquier descripción funcional que trace la frontera entre lo interior y lo exterior en la superficie del cerebro será una descripción funcional según la cual tú y el Gemelo son funcionalmente iguales; pues la organización funcional interna sobreviene en el estado físico interno tanto como sobreviene en él un estado cualitativo.

Pero, los estados funcionales internos del cerebro no pueden captar el contenido de la experiencia, porque este contenido depende de los colores del mundo externo. He aquí una forma de representar esta cuestión. El lente inversor (y también el inversor posretinal) invierten el color sólido 180 grados.[24] Pero de la misma manera se pueden imaginar inversores que inviertan el color sólido en proporciones más pequeñas, por ejemplo, a 90 grados. Imaginemos una serie de clones tuyos (y del Gemelo) cuyos lentes poseen una serie de inversores distintos. Imaginemos que a cada uno de estos clones se les presenta una mancha de color, seleccionada a propósito para producir exactamente la misma señal neuronal en cada uno de los cerebros idénticos. Si su cerebro tiene ante sí una mancha roja, otro tendrá ante sí una mancha amarilla, el tercero una mancha azul, etc. Pero todos tendrán exactamente el mismo contenido cualitativo. Si * es la señal neuronal que estamos considerando como entrada a tu cerebro, entonces * sería producida por el rojo en tu cerebro, por el amarillo en uno de tus clones y por el azul en otro. De suerte que ni * ni ningún estado funcional interno del cerebro que * provocara podría corresponder a un único color del mundo. El contenido intencional de tu experiencia es *verse rojo*, mientras que el de tus clones es *verse amarillo, verse azul*, etc.[25] Todos tus contenidos intencionales son diferentes, pero como

[24] Estoy pensando en una inversión de la rueda de saturación de tonos y de luminosidad constante.

[25] Puesto que sus contenidos cualitativos son los mismos, pero sus contenidos intencionales difieren, *verse rojo* no es un contenido cualitativo. De este

(por hipótesis) sus cerebros son exactamente iguales y como la organización funcional interna sobreviene en el cerebro, **todas** las descripciones funcionales internas serán compartidas por todos ustedes. Así, ninguna descripción funcional puramente interna del cerebro puede captar el contenido intencional de la experiencia.[26]

Lo que acabo de argumentar es que ningún estado funcional interno del cerebro podría ser adecuado para identificarlo con un contenido intencional de la experiencia. Si los contenidos intencionales de la experiencia son estados funcionales, son estados funcionales "de largo alcance" que se extienden hacia el mundo de las cosas y sus colores. Pero, aunque Harman se compromete a reconocer **sólo** estados funcionales de largo alcance, no hay una buena razón por la que el funcionalista no pueda reconocer **tanto** estos papeles funcionales de largo alcance **como** los papeles funcionales de corto alcance que son puramente internos. ¿Por qué no puede el funcionalista identificar los contenidos intencionales con los estados funcionales de largo alcance y el contenido cualitativo con los estados funcionales de corto alcance? El resultado sería una especie de versión "dualista" o de "dos factores" del funcionalismo.[27]

Mi respuesta es que una teoría de dos factores quizá resulte viable, pero el peso de la prueba recae sobre el funcionalista,

modo advertimos que el caso de la Tierra Invertida puede emplearse para desenmascarar la falacia de la intencionalización de los *qualia*, tal como puede hacerse con el caso del espectro invertido.

[26] Lo que marca la diferencia entre el contenido intencional de la experiencia de *verse rojo* y *verse azul* es la conexión entre estas experiencias y los objetos coloreados del mundo. Por lo tanto, cualquier concepción del contenido que deje fuera la conexión entre la experiencia y el mundo no podrá distinguir en forma alguna estos contenidos intencionales distintos. El "contenido estrecho", el aspecto del contenido que está "dentro de la cabeza" es simplemente dicha concepción del contenido, por lo cual no resulta difícil apreciar que el "contenido estrecho" de *verse rojo* y *verse verde* debe ser el mismo. En Block 1987 argumento que una teoría que utilice los papeles funcionales "de largo alcance" postulados por Harman (los que he adoptado en este artículo) es equivalente a una teoría de "dos factores" que emplee papeles funcionales "de corto alcance" junto con un componente referencial.

[27] Véase el escrito citado en la nota anterior (Block 1987) para una discusión de tal teoría de dos factores dentro del contexto de las actitudes proposicionales.

quien tiene que explicarnos cómo son esos estados funcionales de corto alcance. Si carecemos de indicios acerca de cómo serían estos estados funcionales, no tenemos razones para creer en semejante teoría funcional de los *qualia*.[28] De acuerdo con las versiones maquinistas del funcionalismo, la responsabilidad del funcionalista consiste en encontrar una caracterización de los *qualia* en términos del "programa" de la mente. ¡Buena suerte![29]

He concedido que **si** hay una caracterización funcional de los *qualia*, ésta frustrará el Argumento de la Tierra Invertida. Pero esta concesión no le brinda al espectro invertido ninguna superioridad sobre la Tierra Invertida, pues la misma posibilidad frustraría también el argumento del espectro invertido. Se supone que los gemelos del espectro invertido son cualitativamente distintos, pero funcionalmente idénticos. Y se supone que esto demuestra que los estados cualitativos no son estados funcionales. Pero si hay una teoría funcional de los *qualia*, en ella los gemelos del espectro invertido serán, al fin y al cabo, funcionalmente distintos.

Quizá sea posible una teoría fisiológica de los *qualia*. Si la hay, ésta podría funcionalizarse.[30] Sin embargo, tal maniobra cambia el tema dentro del contexto del argumento del espectro invertido. El funcionalismo en ese contexto se encuentra vinculado con el modelo computacional de la mente. La idea consiste más bien en que dos máquinas con distinto *hardware* pueden tener *qualia* diferentes a pesar de la identidad computacional, de que posean la misma estructura computacional y corran los mismos programas. Puesto que la cuestión es si los *qualia* son estados computacionales, no se puede defender legítimamente una teoría computacionalista recurriendo a la propia diferencia en el *hardware*.

[28] También pienso que un argumento de los "*qualia* ausentes" echaría por tierra semejante propuesta, pero no puedo internarme en la cuestión en este lugar.

[29] Adopto el punto de vista de que el pensamiento es funcional, pero que la mejor posibilidad para una teoría de los *qualia* yace en la neurofisiología. Véase Block 1980a.

[30] Como la defendida por Lycan (1987). Describí esta idea como "fisiofuncionalismo" en Block 1980a.

Volvamos ahora a la descripción más plausible como descripción funcional alternativa, cuyo fin es hacerte funcionalmente idéntico al Gemelo, a saber, una descripción que explote el hecho de los lentes inversores. Concentrémonos en la versión en la cual el lente inversor se implanta en el cristalino del ojo. Cabría entonces la objeción de que cuando tú y el Gemelo contemplan sus cielos respectivos, ambos tienen las **mismas** entradas. Los ojos de **ambos** reciben una luz azul, a pesar del hecho de que el Gemelo está mirando un cielo amarillo. Así que, no se produce una inversión funcional.

He estado utilizando una descripción funcional (como lo hace Harman) en la cual las entradas pertinentes a las experiencias como de rojo son los objetos rojos del mundo. La objeción que estoy considerando consiste en que la frontera entre lo interno y lo externo debería correrse hacia dentro, y habría que pensar en las entradas en términos del color de la luz que cae sobre la retina, no en términos de los objetos coloreados. He aquí una primera aproximación a mi respuesta: ofrézcanme una concepción de la frontera entre lo interno y lo externo al sistema, que esté basada en razones independientes, y yo crearé un ejemplo sobre la Tierra Invertida que se ajuste. En este caso, el ajuste está hecho. Se recordará que mencioné una opción a los lentes inversores, la cual me parece físicamente más plausible que la de los lentes, a saber, la de un inversor neuronal colocado detrás de la retina. La retina reacciona de manera bien conocida a tripletes de la longitud de onda de la luz, transformándolas en impulsos en el nervio óptico. No conozco ninguna razón para que, en principio, no fuera posible que un *chip* de silicio en miniatura registrara esos impulsos y los sustituyera por los impulsos transformados.

La sensación es por lo general el producto de la acción del mundo sobre un transductor que registra cierta característica del mundo y emite una señal correspondiente que transmite información al cerebro. El aparato visual registra la luz, el aparato auditivo el sonido, y así con los demás. La manera natural de pensar en las entradas y salidas es en términos de impulsos sobre tales transductores. Con cualquier transductor real, a menudo hay cierta indeterminación sobre el límite exacto donde empieza o termina la transducción. Se podría pensar en el cris-

talino del ojo como el transductor visual o, como alternativa, se podría considerar que el transductor visual termina en los bastones y conos o en una de las otras capas de la retina, o en el nervio óptico, o en el núcleo geniculado lateral. Yo podría ubicar mi inversor en cualquiera de esos sitios. Sin embargo, supongo que podría descubrirse la existencia de un "sensor visual" en el cerebro con un límite natural definido, de suerte que cualquier cambio que se produce en su interior altera los *qualia*, y cualquier cambio en su exterior no los altera (excepto al afectar sus entradas). En tal caso, la selección del límite del sensor como la frontera entre lo interno y lo externo exigiría que el inversor se localizara en un sitio donde se alterasen los *qualia*. A esta posibilidad respondo con la observación, que también hice en la sección sobre la falacia de la intencionalización de los *qualia*, de que podría haber un tipo de persona, muy semejante a nosotros en los demás aspectos, que tuviera visión y sensaciones de colores, pero cuyo sensor careciese de límites definidos. Luego, puedo aplicar mis experimentos mentales a un tipo de ser (tal vez hipotético, o tal vez a nosotros mismos) que tiene experiencias de color, pero que carece de un sensor claramente delimitado.

3. *La conciencia de las propiedades intrínsecas*

Hasta ahora he estado refutando las objeciones de Harman al espectro invertido y exponiendo mi propio argumento en términos de la Tierra Invertida para la aceptación de la distinción entre lo cualitativo y lo intencional, y la falsedad de la teoría funcionalista de los *qualia*. Sin embargo, hay una consideración planteada por Harman que todavía no he mencionado. Dice que la distinción entre lo cualitativo y lo intencional depende de la idea de que somos conscientes de las propiedades intrínsecas de nuestra experiencia.

Dedica más de la mitad de su artículo a rebatir esta idea. De modo que con el fin de examinar cabalmente su objeción a la hipótesis del espectro invertido, tendremos que considerar su objeción a la idea de que somos conscientes de las características intrínsecas de nuestra experiencia.

El argumento básico de Harman es, hasta donde alcanzo a ver, una invitación —entre todas las cosas— a la introspección. Afirma que la idea de que somos conscientes de la "pintura mental" contradice "la experiencia visual ordinaria". He aquí el pasaje principal donde se expone el argumento:

> Cuando Eloísa ve un árbol ante ella, los colores que experimenta son experimentados todos como características del árbol y de sus alrededores. No experimenta ninguno como característica intrínseca de su experiencia. Tampoco experimenta ninguna de las características de otras cosas como características intrínsecas de su experiencia. Y esto es verdadero también en tu caso. [...] Contempla un árbol e intenta fijar la atención en las características intrínsecas de tu experiencia visual. Predigo que descubrirás que las únicas características sobre las cuales fijar la atención ahí serán las características del árbol presentado [...].

Desde mi punto de vista, esta invitación de Harman a la introspección constituye un error de método filosófico. Cuando contemplo mi pared azul, **pienso** que, además de ser consciente del color, también puedo tomar conciencia de lo que es para mí ser consciente del color. Estoy seguro de que habrá quienes discrepen de mí. Lo único en que todos debemos estar de acuerdo es en que ésta no es manera de hacer filosofía. Nuestras intuiciones pueden darnos indicios sobre nuestros conceptos y nuestras experiencias mientras las usemos correctamente. Un principio es: **obtenga intuiciones simples sobre casos complejos, en vez de intuiciones complejas sobre casos simples**. Es fácil contemplar una pared azul, pero no lo es (quizá no resulta posible) contestar, sólo sobre la base de la introspección, a la cuestión altamente teórica de si al hacerlo soy consciente de las propiedades intrínsecas de mi experiencia. La finalidad de las complejas historias de ciencia ficción descritas anteriormente consiste en producir casos complejos sobre los cuales uno puede consultar intuiciones simples. Harman espera rebatir el espectro invertido recurriendo al juicio introspectivo de que no somos conscientes de las características intrínsecas de nuestra experiencia. Creo que toma el camino equivocado. Si argumentos como el que expuse nos convencen de que es posible que el

contenido cualitativo siga siendo el mismo mientras que el contenido intencional se invierte (o al revés), esto decidirá entonces la cuestión de si somos o no conscientes de las características intrínsecas de nuestra experiencia. O en todo caso, decidirá la cuestión que está realmente en juego: si **hay** o no características mentales intrínsecas de nuestra experiencia.

Finalmente, pasaré a la crítica que hace Harman a Jackson. Expongamos la cuestión tomando a la María de Jackson, la científica que conoce todos los hechos físicos y funcionales acerca de la percepción de color que pueden aprenderse en un entorno blanco y negro. El problema que plantea Jackson es el siguiente: ¿Adquiere María un nuevo conocimiento al ver rojo? El realista de los *qualia* dice que adquiere conocimiento genuino **de que**, información genuina y nueva, a saber, el conocimiento de que así es cómo es ver rojo. Lewis, Nemirow, Shoemaker y otros críticos han señalado que lo que María adquiere es algo más semejante al conocimiento **cómo**, nuevas habilidades en vez de nueva información. La posición de Harman es que María **sí** adquiere realmente información nueva, genuino conocimiento de que –y en este punto parece (engañosamente) incorporar la intuición del realista de los *qualia*–, pero continúa diciendo que el funcionalismo no queda refutado porque María no conocía todos los hechos funcionales antes de ver rojo. Harman afirma que el conocimiento de todos los hechos funcionales exige la posesión del concepto completo de rojo y que esto es cuestión de encontrarse en cierto estado funcional. Más específicamente, el concepto completo de rojo debe ser producido de manera natural e inmediata por las cosas rojas y debe desempeñar cierto papel en el pensamiento y en la acción con respecto a las cosas rojas. Mientras siga encerrada en el cuarto en blanco y negro, María no se encuentra en un estado producido de manera natural e inmediata por las cosas rojas, y que desempeñe el papel pertinente en el pensamiento y la acción. Así María carecía de ese estado funcional antes de ver rojo; por ende, no sabía todos los hechos funcionales. Cuando ve una cosa roja por primera vez adquiere el concepto de rojo, y con éste los hechos funcionales de los que carecía al carecer de ese concepto. Además, ahora es posible decir que sabe que la sangre es roja, que las manzanas son rojas, etc., mientras que antes podía pronunciar

esas palabras pero, al faltarle el concepto de rojo, no poseía realmente el conocimiento al respecto.

En suma, Jackson afirma que María conoce todos los hechos funcionales antes de ver algo rojo, pero aun así aprende un nuevo hecho, de modo que los hechos funcionales no incluyen todos los hechos. Algunos de los críticos de Jackson niegan que María aprenda un nuevo hecho. Harman, en cambio, admite que María aprende nuevos hechos al ver rojo, pero niega que María supiera todos los hechos funcionales antes de ver algo rojo. Dice que adquiere también un nuevo hecho funcional al adquirir el concepto mismo de rojo.

Si Harman está en lo correcto en todo esto —por mor del argumento, supongamos que lo está—, no ha ofrecido ningún argumento en contra del realismo de los *qualia*; más bien, ha mostrado que el argumento de Jackson no es muy útil como *locus* de controversia entre el realista y el escéptico de los *qualia*. El conocimiento del que carece María mientras está en el cuarto en blanco y negro, según muestra Harman, tiene poco que ver con los verdaderos puntos en discusión acerca de los *qualia*. No es sino hasta que María ve verde por primera vez cuando puede tener el concepto completo de verde, y por lo tanto, no puede tener el conocimiento de que algunas aceitunas son verdes. En verdad, ni siquiera puede tener la falsa creencia de que todas las aceitunas son verdes. Este hecho tiene poco que ver con el desacuerdo entre el escéptico y el realista de los *qualia*.

Si se acepta la afirmación de Harman de que María no puede tener creencias acerca de los colores de las cosas hasta que salga del cuarto en blanco y negro, el verdadero punto de discusión entre el realista y el escéptico de los *qualia* acerca de María es entonces que el realista de los *qualia* dice que María adquiere **dos** tipos de conocimiento *de que* cuando ve los colores por primera vez, mientras que el escéptico de los *qualia* asegura que sólo adquiere uno. Todos pueden estar de acuerdo (suponiendo que Harman está en lo correcto acerca de los conceptos) en que María adquiere un conocimiento que involucra contenidos **intencionales**. Por ejemplo, adquiere el concepto de verde y, como ya ha leído la *Encyclopedia Brittanica*, adquiere el conocimiento de que las aceitunas son verdes. Sin embargo, el realista de los *qualia* dice que María adquiere un conocimiento

que involucra contenidos cualitativos además de los contenidos intencionales. Éste es el meollo de la cuestión y, a este respecto, la respuesta de Harman a Jackson permanece muda.

[*Traducción de Ana Isabel Stellino*]

BIBLIOGRAFÍA

Block, N., 1987, "Functional Role and Truth Conditions", *Proceedings of the Aristotelian Society*, supl. 61, pp. 157–181.

———, 1980a, "Troubles with Functionalism", en Block 1980b, pp. 268–305.

———, 1980b, *Readings in Philosophy of Psychology*, vol. 1, Harvard University Press, Cambridge.

———, 1980c, "Are Absent Qualia Impossible?", *Philosophical Review*, vol. 89, pp. 257–274.

Churchland, P., 1979, *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge.

Dennett, D., 1992, "Quining Qualia", en Marcel y Bisiach 1988, pp. 619–642. [Incluido en esta antología, pp. 213–262.]

———, 1979, "On the Absence of Phenomenology", en Gustafson y Tapscott 1979, pp. 99–103.

———, 1978a, *Brainstorms: Philosophical Essays on Mind and Psychology*, MIT Press, Cambridge, Mass.

———, 1978b, "Toward a Cognitive Theory of Consciousness", en Dennett 1978a, pp. 149–173.

———, 1978c, "Why You Can't Make a Computer that Feels Pain", *Synthese*, vol. 38, no. 3, pp. 415–449. (Reimpreso en Dennett 1978a, pp. 190–229.)

Gustafson, D. y B. Tapscott (comps.), 1979, *Body, Mind and Method: Essays in Honor of Virgil Aldrich*, Reidel, Dordrecht.

Hardin, C.L., 1988, *Color for Philosophers: Unweaving the Rainbow*, Hackett, Indianápolis.

Harman, G., 1982, "Conceptual Role Semantics", *Notre Dame Journal of Formal Logic*, vol. 23, no. 2, pp. 242–256.

Harrison, B., 1973, *Form and Content*, Oxford University Press, Oxford.

Haugeland, J., 1978, "The Nature and Plausibility of Cognitivism", en *The Behavioral and Brain Sciences*, vol. 1, no. 2, pp. 215–226.

Jackson, F., 1977, *Perception*, Cambridge University Press, Cambridge.

Lycan, W., 1987, *Consciousness*, MIT Press, Cambridge, Mass.
———, 1973, "Inverted Spectrum", *Ratio*, vol. 15, pp. 315–319.
Marcel, A. y E. Bisiach (comps.), 1988, *Consciousness in Contemporary Science*, Oxford University Press, Oxford.
McClamrock, R., 1984, tesis de doctorado, Massachusetts Institute of Technology, Cambridge, Mass.
McGinn, C., 1983, *The Subjective View*, Oxford University Press, Oxford.
Peacocke, C., 1983, *Sense and Content*, Oxford University Press, Oxford.
Rey, G., inédito, *Mind without Consciousness: A Discrepancy between Explanatory and Moral Psychology*.*
Searle, J.R., 1992, *The Rediscovery of the Mind*, MIT Press, Cambridge, Mass. [Versión en castellano: *El redescubrimiento de la mente*, trad. L.M. Valdés Villanueva, Crítica, Barcelona, 1996.]
Shoemaker, S., 1984, *Identity, Cause and Mind: Philosophical Essays*, Cambridge University Press, Cambridge.
———, 1981, "Absent Qualia Are Impossible? –a Reply to Block", *Philosophical Review*, vol. 90, no. 4, pp. 581–599.
———, 1982, "The Inverted Spectrum", *Journal of Philosophy*, vol. 79, no. 7, pp. 357–381.
———, 1975, "Functionalism and Qualia", *Philosophical Studies*, vol. 27, pp. 291–315. (Reimpreso en Block 1980b, pp. 251–267.) [Incluido en esta antología, pp. 183–212.]
Tye, M., 1985, Reseña de *Sense and Content* de C. Peacocke, *Canadian Philosophical Reviews*, vol. 5, pp. 173–175.

* Aunque estaba prevista su publicación en MIT Press, Rey la retiró y nunca apareció impresa. Él mismo hace la aclaración en *Contemporary Philosophy of Mind*, Blackwell, Oxford, 1997.

UNA TEORÍA REPRESENTACIONAL DEL DOLOR Y DE SU CARÁCTER FENOMÉNICO*

MICHAEL TYE

El supuesto fundamental de la psicología cognoscitiva consiste en que la mente es un *sistema representacional* que media entre las entradas sensoriales y las salidas conductuales. La tarea primordial del psicólogo cognoscitivo es explicar de qué manera operan las diversas capacidades cognoscitivas haciendo referencia a la estructura de las partes distintivas de este sistema representacional. Las explicaciones ofrecidas son tanto funcionales como descomposicionales: descomponen las capacidades pertinentes en sus elementos representacionales básicos y muestran cómo esos elementos funcionan conjuntamente para producir dichas capacidades. Las teorías se evalúan por el grado de exactitud con que explican el comportamiento observado en los experimentos psicológicos y, en su nivel más bajo, por el grado en que se adecúan al conocimiento que la neurofisiología ha recabado sobre las bases físicas de las capacidades.

Los filósofos han asumido generalmente que el dolor no puede caer dentro del dominio de la psicología cognoscitiva. Los dolores, según se ha supuesto, no se asemejan a imágenes, recuerdos o percepciones visuales: carecen de contenido representacional. Por lo tanto, no puede haber una explicación de la clase deseada. Para comprender las diversas facetas del dolor,

* Quiero agradecer a Ned Block, Gabriel Segal y, especialmente, a Sydney Shoemaker sus útiles comentarios.

necesitamos buscar en otro sitio, quizá en el reino de la neurofisiología. La psicología cognoscitiva no puede ayudarnos. Esto, según creo ahora, es un error: los dolores *sí* poseen contenido representacional. En consecuencia, la concepción de que el dolor no constituye un objeto de estudio apropiado para la psicología cognoscitiva carece de buenos fundamentos.

Empiezo la discusión con una vieja objeción a la teoría de identidad caso con relación a las imágenes residuales, y con una respuesta moderna a ésta, que ha encontrado amplia aceptación. Sostengo que esta respuesta, tal como se halla formulada, resulta insatisfactoria. Sin embargo, si se le hace un ajuste crucial es, a mi parecer, defendible, y tiene ramificaciones que contribuirán a nuestra comprensión del dolor. En particular, apunta a la conclusión de que los dolores tienen contenido representacional, como también lo indica por lo menos una de las otras facetas de nuestra concepción cotidiana del dolor. En la tercera sección del artículo, considero el problema de qué clases de representaciones se supone, más plausiblemente, que son los dolores. ¿Son oraciones de un lenguaje interno, como las creencias y los deseos lo son en la concepción computacional? ¿O son representaciones de una clase diferente? Sugiero que un enfoque oracional resulta difícil de conciliar con algunos de los datos neurofisiológicos sobre el dolor, y desarrollo una propuesta alternativa. De acuerdo con ésta, los dolores serían representaciones de la misma clase general que las imágenes mentales: ordenamientos a los que se adjuntan etiquetas descriptivas. De suerte que el dolor, insisto, constituye un adecuado objeto de estudio para la psicología cognoscitiva. En la sección final, retomo algunas cuestiones relativas al carácter fenoménico del dolor.

I

En los años cincuenta, Smart (1959) planteó la siguiente objeción a la teoría de la identidad de las sensaciones: las imágenes residuales son en ocasiones anaranjado-amarillentas; los procesos cerebrales no pueden ser anaranjado-amarillentos. Por lo tanto, las imágenes residuales no son procesos cerebrales. La respuesta que el propio Smart dio a esta objeción fue negar

la existencia de las imágenes residuales, que en realidad serían, según Smart, sólo experiencias de *tener* imágenes residuales, las cuales no son en sí mismas anaranjado-amarillentas.

Existe otra respuesta menos radical en beneficio de la teoría de la identidad. En esta respuesta me quiero centrar aquí. ¿Por qué no decir que, al adjudicar predicados de color a las imágenes no les estamos atribuyendo las mismas propiedades que atribuimos a los objetos externos mediante el uso del lenguaje de color? De este modo, las imágenes residuales no son literalmente verdes o azules de la misma manera en que el pasto o el cielo tienen una u otra de estas características. Ahora ya no es evidentemente verdadero que los procesos cerebrales no puedan ser anaranjado-amarillentos en el sentido pertinente.

El problema obvio que afronta esta respuesta es explicar cómo puede siquiera aplicarse el vocabulario del color a las imágenes residuales, dado que no poseen realmente los colores apropiados. Una solución, propuesta por Ned Block, consiste en decir que las palabras relativas a los colores se utilizan elípticamente para expresiones como "representar el azul verdadero", "representar el verde verdadero", y demás, en conexión con las imágenes en general (Block 1983, esp. la p. 518). Desde mi punto de vista, esta solución tiene varias virtudes importantes.[1] En primer lugar, los procesos cerebrales pueden sin duda representar colores. Por lo tanto, la teoría de la identidad ya no está amenazada. En segundo lugar, según lo ha observado Block (Block 1983, pp. 516–517), términos como "alto" y "agudo" se aplican de manera convencional y directa a las lecturas osciloscópicas utilizadas en relación con las representaciones gráficas de los sonidos. En este contexto, es evidente que estos términos no nombran los sonidos reales obtenidos por las lecturas. Una posibilidad es, pues, que seleccionen propiedades representacionales, como representar lo fuerte y representar lo agudo. De ser así, entonces ya existe un uso establecido de términos que coincide con el que supuestamente se da en el caso de los términos referentes a los colores y las imágenes residuales.

[1] Estas virtudes me llevaron a aceptar la propuesta hasta hace muy poco tiempo. Véase Tye 1991.

Hay, no obstante, una seria dificultad. Las imágenes mentales no son literalmente cuadradas, como tampoco son literalmente azules. Por lo tanto, extendiendo la propuesta anterior a la figura, tenemos que una imagen residual cuadrada y azul no es sino una imagen residual que representa lo cuadrado y que también representa lo azul. Pero, en el nivel intuitivo, esto parece demasiado endeble. Sin duda, una imagen cuadrada azul no puede representar cosas distintas como lo cuadrado *y* lo azul. Desafortunadamente, la propuesta citada no contiene nada que descarte esto. "Azul", pues, cuando se aplica a las imágenes residuales, no significa "representar lo azul". Lo mismo ocurre con "cuadrado".

Esta dificultad no es exclusiva de las imágenes. Precisamente el mismo problema puede plantearse con relación a las lecturas osciloscópicas. Sugiero que la vía de escape consiste en darnos cuenta de que no hay nada elíptico en el significado de términos como "azul" o "alto" en los contextos mencionados. Son, más bien, los contextos mismos los que necesitan ser examinados más detenidamente. Permítaseme explicar esto.

Los contextos "Espera *F*" y "Alucina *G*" son típicamente intensionales. Así, puedo esperar la vida eterna y alucinar un elefante rosa, aun cuando tales cosas no existan. De modo semejante, puedo esperar la vida eterna sin esperar el aburrimiento eterno, aunque ambas cosas sean en realidad lo mismo. Parece evidente que los términos que sustituyen a "*F*" y "*G*" en estos contextos conservan su significado habitual. Las peculiaridades que acabo de mencionar se deben al hecho de que esperar y alucinar son estados representacionales, y al carácter especial de la representación misma.

Ahora bien, precisamente las mismas peculiaridades están presentes en el caso del contexto "una imagen *F*", donde "*F*" es un término de color o de forma. De esta manera, en un mundo en el cual no hay nada realmente triangular, puedo tener, no obstante, una imagen triangular. Asimismo, si tengo una imagen roja, intuitivamente no se sigue que tengo una imagen del color de la mayoría de los camiones de bomberos, aun cuando la mayoría de los camiones de bomberos sean rojos. Sugiero que

la explicación es más directa: una imagen *F* es una imagen que *representa que algo* es *F*.[2]

De manera similar, una imagen *F*, *G* es una imagen *que representa que algo* es a la vez *F* y *G*. Mi sugerencia es, pues, que no hay nada elíptico o peculiar en el significado de los términos "*F*" y "*G*" en el contexto "Una imagen *F*, *G*". Más bien, el contexto mismo es intensional y posee una estructura lógica que refleja el carácter representacional de las imágenes en general.

Cabría preguntar, no obstante, por qué *decimos* que la imagen misma es *F* y *G*, por ejemplo, azul y cuadrada. Sugiero que esto forma parte de un uso mucho más amplio. Frecuentemente, cuando hablamos de representaciones, tanto mentales como no mentales, en la ciencia y en la vida cotidiana, nos ahorramos palabras hablando como si las representaciones mismas tuvieran las propiedades de las cosas que representan. En tales casos, al decir que una representación es *F*, lo que queremos expresar es que representa que algo es *F*. De este modo, cuando se dice que una lectura osciloscópica dada es alta y aguda, lo que estamos afirmando es que alto y agudo son rasgos que la lectura representa como propias de un sonido. "Alto" y "agudo" significan aquí lo que normalmente significan. El contexto mismo es intensional.

La propuesta anterior resuelve el problema de la imagen cuadrada azul. En este caso la imagen representa que algo es a la vez azul y cuadrado, no simplemente que algo es azul *y* que algo es cuadrado.[3]

[2] Podría agregar que hay una posible lectura *de re* de este contexto como sigue: la cualidad *F* es tal que una imagen *F* representa que algo *la* posee. Ahora bien, a partir de "Tengo una imagen roja" y "Rojo es el color de la mayoría de los camiones de bomberos", *puede* inferirse "Tengo una imagen del color de la mayoría de los camiones de bomberos".

[3] Podría añadir que, en mi opinión, una condición necesaria para cualquier imagen que represente que algo es a la vez *F* y *G* es que represente que algo es *F*. Por lo tanto, si tengo una imagen *F*, *G*, debo tener una imagen *F*. El argumento para la premisa resulta, en este caso, claro: cuando tengo una imagen azul cuadrada, experimento el azul *como* un rasgo de una cosa u otra, un rasgo coinstanciado con lo cuadrado. *Lo que* experimento es, en parte, *que* algo es azul. En consecuencia, mi imagen representa, en parte, que algo es azul.

Podría agregar que la afirmación de que las imágenes residuales son representacionales no implica ni presupone que las criaturas no puedan tener imágenes residuales a menos que también tengan los conceptos apropiados (por lo menos, según suele entenderse la palabra "concepto"). Tener el concepto *F* exige, de acuerdo con ciertas explicaciones, tener la capacidad de usar el término lingüístico "*F*" de manera correcta. De acuerdo con otras explicaciones, la posesión de un concepto requiere la capacidad para representarse en el pensamiento y en la creencia que algo cae bajo ese concepto. Pero las imágenes residuales, como otras sensaciones perceptuales, no son ellas mismas pensamientos o creencias; y sin duda no exigen un lenguaje público. Son, si se quiere, representaciones no doxásticas o no conceptuales.

La idea general que tengo es, muy concisamente, la siguiente: los procesos que operan con base en estímulos proximales generan ciertas clases de representaciones visuales a través de procedimientos mecánicos. Las categorizaciones de las diversas clases se van produciendo durante el mismo proceso. Las sensaciones visuales son representaciones que forman las salidas de este temprano procesamiento modular, y están prontas a producir respuestas conceptuales a través de la acción de una u otra clase de procesamiento cognoscitivo de alto nivel. De ese modo, las sensaciones visuales alimentan el sistema conceptual, sin que ellas mismas formen parte de este sistema.

Al admitir que las imágenes residuales son representaciones no conceptuales, no estoy concediendo que carezcan realmente de contenido *intencional*, que se trata de representaciones de tipo no intencional. En mi opinión, la intencionalidad no requiere conceptos. Según mi uso del término "intencionalidad", los rasgos cruciales son la representación (y de aquí la posibilidad de una representación *errónea*) junto con la posibilidad de sustituciones fallidas en los contextos lingüísticos relacionados. Estos rasgos, como he argumentado, están presentes en el caso de las imágenes residuales. Por lo tanto, las imágenes residuales son intencionales. Los filósofos que quieren insistir en que no puede haber intencionalidad genuina sin conceptos (o que la intencionalidad se restringe a la parte ejecutiva central) tienen derecho a utilizar el término en este sentido. Pero sugiero que

cualquier desacuerdo al respecto es puramente verbal. No hay nada sustancial implicado en cuál es el uso que se adopta.

Estamos ahora preparados para pasar al caso del dolor.

II

Se supone frecuentemente que los términos aplicados al dolor y que también se aplican a los objetos físicos no poseen su significado ordinario. Ned Block, quien asume este punto de vista, señala lo siguiente:

> Tenemos ciertos motivos para pensar que hay una diferencia de significado sistemática entre algunos predicados que se aplican a los objetos físicos y los mismos predicados cuando se aplican a particulares mentales. Consideremos un ejemplo en lenguaje no figurado: el predicado _____ en _____. Este predicado aparece en ambas premisas y en la conclusión del siguiente argumento:
>
> El dolor está en la punta de mi dedo.
> La punta de mi dedo está en mi boca.
> Por lo tanto, el dolor está en mi boca.
>
> Este argumento es válido para el "en" de limitación espacial..., ya que en este sentido "en" es transitivo. Pero supongamos que las dos premisas son verdaderas en cuanto a sus significados *ordinarios* [...]. La conclusión, evidentemente, no se sigue, de suerte que debemos concluir que "en" no ha sido usado en el sentido de limitación espacial en las tres oraciones. Parece sin duda factible que el "en" que se aplica para ubicar dolores difiere sistemáticamente en cuanto a su significado de la acepción convencional de limitación espacial. (Block 1983, p. 517)

Esto me parece totalmente erróneo. No hay mayor motivo para adoptar la extraña posición de que "en" no significa encerramiento espacial en relación con el dolor que para decir que "anaranjado", en relación con las imágenes, posee un significado especial. Al derrumbarse esta última concepción, la primera se tambalea. Y la inferencia que Block cita *no* establece su afirmación. Para entender esto, consideremos la siguiente inferencia:

Quiero estar en el palacio municipal.

El palacio municipal está en un gueto.
Por lo tanto, quiero estar en un gueto.

El término "en" posee el mismo significado en las dos premisas y en la conclusión. Pero el argumento carece de validez: quizá quiera yo estar en el palacio municipal para, digamos, escuchar un discurso particular, sin por ello querer estar en un gueto. Sugiero que esto también es cierto para el caso del ejemplo de Block, y la explicación es la misma. Tanto en la primera premisa como en la conclusión, el término "en" aparece dentro de un contexto intensional. Así como cuando decimos que una imagen es azul, estamos diciendo que representa que algo es azul, también cuando decimos que un dolor está en la punta de mi dedo, estamos diciendo que representa que algo está en la punta de mi dedo.

Quizá valga la pena observar aquí que la invalidez de la inferencia concerniente al dolor no tiene nada que ver con el hecho de que la boca es una cavidad de cierto tipo y que por lo tanto es un elemento cuya condición ontológica misma podría ser cuestionada. Si tengo un dolor en la punta del dedo y me hago entonces una pequeña incisión en la pierna donde luego coloco mi dedo, de ello no se sigue aún que tengo un dolor en la pierna. Supongamos, por ejemplo, que mi pierna ha sido anestesiada. En este caso, siento dolor en el dedo, pero no en la pierna.

Tampoco nos ayuda decir que lo que realmente muestra el fracaso de la inferencia es que los dolores mismos resultan ontológicamente dudosos. Pues aun cuando fuera cierto que no hay dolores, sino sólo personas doloridas, esto continúa sin explicarnos por qué falla la inferencia. Después de todo, si "en" significa adentro, y yo tengo la punta del dedo dolorida, también tengo entonces la boca dolorida, suponiendo que la punta de mi dedo esté en mi boca.

El hecho de que existe un carácter intensional oculto en las oraciones sobre la ubicación del dolor se confirma por la manera en que hablamos de dolores en miembros amputados. Admitimos como verdadera la afirmación de que las personas sufren dolores en miembros que ya no existen. ¿Cómo puede ser esto? La respuesta es que uno puede sentir dolor en la pier-

na izquierda aun cuando ya no tenga la pierna izquierda, tal como se puede buscar la Fuente de la Juventud. También en este caso el contexto es intensional: específicamente, uno tiene un dolor que representa que algo está en la pierna izquierda.[4]

Desde luego, existe la tentación de decir que, si uno no tiene la pierna izquierda, no puede sentir realmente dolor en ella. Pero esto no entraña ninguna dificultad para mi propuesta. Pues hay una lectura *de re* del contexto, a saber, que tener un dolor en la pierna izquierda implica que la pierna izquierda es tal que se tiene un dolor en *ella*. Ahora sí se requiere una pierna izquierda.

Pero un dolor en la pierna, ¿no representa más que el mero hecho de que algo está en la pierna? Para responder a esta pregunta, es necesario hacer otras observaciones generales acerca del dolor. Tener un dolor es sentir un dolor, y sentir un dolor es experimentar dolor. De este modo, si tengo un dolor, tengo una experiencia caso de cierto tipo. Esta experiencia caso es, sugiero, el dolor particular que tengo.

La identificación de los dolores con ejemplares de experiencias caso explica por qué ninguna otra persona podría sentir el dolor *particular* que estoy sintiendo e, incluso, por qué los dolores pertenecen necesariamente a *una* u otra persona. Las experiencias caso son sucesos (en el sentido amplio, que incluye los casos de estados) y, en general, los sucesos se individúan en parte a través de los objetos que los padecen. Así, la risa no puede existir sin dueño, y tampoco los gritos. Lo mismo ocurre con los asesinatos, los nacimientos y las explosiones. En cada uno de estos casos, debe haber un sujeto para el suceso: un asesino, una criatura que nace, un objeto que explota. Además, el sujeto está esencialmente relacionado con el suceso. Nadie más puede morir mi muerte o reír mi risa, aunque, desde luego, es cierto que otra persona puede experimentar una muerte o una risa que cualitativamente es muy semejante a la mía.

Ahora bien, las experiencias del dolor, si se encuentran en algún lado es en la cabeza. Pero en el caso de un dolor en la

[4] El dolor en miembros amputados muestra que los dolores no implican esencialmente relaciones entre las personas y las partes de su cuerpo. Me parece que ésta es una objeción decisiva a la concepción relacional presentada en Aune 1967, p. 130.

pierna, lo que localiza la experiencia del dolor, cuando todo está funcionando normalmente, es un daño del tejido en la pierna. Por lo tanto, sugiero que un dolor en la pierna es un caso de experiencia sensorial que representa que algo en la pierna está dañado.[5]

Hasta aquí no he dicho nada que aborde directamente la cuestión de lo doloroso de los dolores. ¿Cómo ha de explicarse esta característica de los dolores en la propuesta que se expuso líneas atrás? Para comenzar, debe observarse que a menudo hablamos de las lesiones corporales calificándolas de dolorosas. Cuando se dice que un corte en el dedo, una quemadura o una contusión son dolorosos o duelen, lo que quiere decirse es (en términos generales) que está *causando* una sensación, a saber, la sensación misma que sufre una persona, y que tal sensación provoca un desagrado inmediato junto con la ansiedad, o la preocupación, por el estado de la región del cuerpo donde se siente que el trastorno está localizado.

Desde luego, los dolores no causan normalmente por sí solos sensaciones que provoquen desagrado: *son* esas sensaciones, por lo menos en los casos típicos. En consecuencia, los dolores no son dolorosos en el sentido antes mencionado. Pero lo son en un sentido ligeramente más débil: provocan típicamente las reacciones *cognoscitivas* que acabo de describir.[6] Además, cuando examinamos introspectivamente nuestros dolores, tomamos conciencia de sus contenidos sensoriales como dolorosos. Ésta es la razón de que, si tengo un dolor en la pierna, tome yo conciencia intuitivamente de que algo en mi pierna es doloroso (y no en mi cabeza, que es donde, en mi opinión, se encuentra la experiencia misma). Mi dolor representa una lesión en mi

[5] En mi opinión, aquí la representación del daño es no conceptual. Por lo tanto, puedo sin duda ver que mi pierna está lesionada sin sentir allí ningún dolor. En el caso de "ver que", mi estado entraña una creencia acerca del daño (y en consecuencia el ejercicio de conceptos).

[6] Según los investigadores del dolor, las personas a quienes se les ha practicado una lobotomía prefrontal, u otros tratamientos, suelen informar que sienten dolor pero que no les preocupa o que realmente no les molesta (véase Melzack 1961, 1973). Estos informes, aun si se toman al pie de la letra, son compatibles con la propuesta del texto. Pues es evidente que tales casos son anormales.

pierna, y clasifico entonces cognoscitivamente esa lesión como dolorosa (aplicando el concepto de *doloroso* en la introspección).

De este modo, en circunstancias normales, una persona que tiene un dolor en la pierna y que informa que algo en su pierna es doloroso no se encuentra bajo ningún tipo de ilusión. Pero es diferente la situación de un hombre que le informa a su médico que tiene un dolor en el brazo izquierdo si se descubre que la verdadera causa de su dolor reside en el corazón. Este hombre tiene un dolor en el brazo izquierdo —sufre una experiencia sensorial que le representa la lesión en ese sitio—, pero no ocurre realmente nada malo *en su brazo izquierdo*. Lo doloroso es un trastorno que está teniendo lugar en su corazón.

Hay una objeción que amerita mencionarse aquí. Se dirá quizá que una persona que experimenta dolor en cierta parte del cuerpo no necesita estar consciente *de que* existe tal o cual daño en el tejido de la parte en cuestión. Tal propuesta es demasiado compleja para encajar en la fenomenología de las experiencias del dolor (o de muchas de esas experiencias).

No es una objeción convincente. Los dolores, desde mi punto de vista, son representaciones *sensoriales* de una lesión tisular. Para sentir un dolor, no se necesita tener los recursos para conceptualizar lo que el dolor representa (véase la nota 5). No es preciso poder decir o pensar que se está produciendo tal o cual daño tisular. Aun así, el contenido de la representación sensorial pertinente es lo que confiere al dolor su carácter fenoménico.

En mi opinión, la explicación intencionalista que he propuesto puede extenderse a dolores de tipos más específicos. Una punzada es un dolor que representa un trastorno breve y ligero. Un dolor palpitante es el que representa un trastorno de pulsaciones rápidas. Otros dolores [*aches*] representan trastornos que se producen *dentro* del cuerpo, más que en la superficie. Estos trastornos se representan como si tuvieran volumen, como si comenzaran y finalizaran gradualmente, como si su gravedad fuera en aumento para luego desaparecer poco a poco (*cfr.* Armstrong 1962). Los volúmenes así representados no se representan como precisos o claramente demarcados. Es por esto que estos dolores no parecen tener una ubicación exacta, al contrario, por ejemplo, de los pinchazos. Un dolor punzante es

el que representa un daño repentino en una región particular del cuerpo perfectamente definida. Esta región se representa con volumen (en vez de ser bidimensional), con la forma de algo afilado y puntiagudo (como un puñal).[7] En el caso de un pinchazo, el daño pertinente se representa con un comienzo y un final repentinos en la superficie o inmediatamente debajo de ella, y cubre una zona muy pequeña. Un dolor lacerante es el que representa un daño que implica el estiramiento de las partes internas del cuerpo (por ejemplo, los músculos).

En cada uno de estos casos, el sujeto del dolor experimenta una representación sensorial de cierta clase de trastorno corporal. Los trastornos varían según los dolores. Consideremos, por ejemplo, un pinchazo en la pierna. En este caso, parece fenomenológicamente innegable que el pinchazo se experimenta *como* un rasgo que se instancia en la pierna, y no como un rasgo intrínseco de la experiencia misma. Lo que se experimenta como lo que se pincha forma parte de la superficie de la pierna. Esto queda perfectamente explicado por la propuesta desarrollada antes. También debe observarse que, como los pinchazos no representan alfileres, mi explicación no conlleva la inverosímil consecuencia de que las criaturas que viven en mundos sin alfileres no pueden tener sensaciones de pinchazos o que en estos mundos las criaturas que sufren tales sensaciones se representan erróneamente lo que les sucede.

Mi conclusión es que los dolores, como las imágenes residuales, poseen contenido representacional. Pero, al contrario de las imágenes, tienen una localización en el cuerpo (en el sentido representacional que he elucidado).[8] Por lo tanto, aunque los dolores están constituidos, en mi opinión, por procesos físicos

[7] No pretendo sugerir aquí que no se puede sentir un dolor punzante a menos que se tenga el concepto de un puñal. Los dolores, repito, son representaciones sensoriales no conceptuales.

[8] Niego que los llamados dolores psicológicos, como el dolor del remordimiento o de la vergüenza, sean verdaderos dolores. Me parece admisible sostener que tales estados se etiquetan como "dolores" debido a que, al igual que ocurre con los dolores (normales), las personas sienten aversión por ellos. Pero sugiero que dicho uso del término "dolor" es metafórico o analógico. Esto no equivale a negar, desde luego, que los dolores reales puedan tener causas psicológicas. Ciertamente, la vergüenza puede *causar* un dolor de ardor en el rostro (véase Stephens y Graham 1987, p. 413).

que se producen en la cabeza, también es cierta la afirmación de que pueden tener lugar en cualquier parte del cuerpo.[9]

III

La hipótesis sobre el lenguaje del pensamiento es una hipótesis empírica sobre la manera en que los contenidos representacionales de los estados mentales se codifican de hecho en la cabeza. No es un análisis filosófico *a priori*. Por lo tanto, no pretende cubrir los contenidos de los estados mentales de todas las criaturas reales y posibles. En su forma más general, se ocupa de la codificación de *todos* los contenidos mentales reales. La tesis básica, que emana del modelo computacional de la mente, consiste en que tales contenidos están codificados en estructuras simbólicas expresadas en un lenguaje interno.

En el caso de lo que ha dado en llamarse actitudes proposicionales —esto es, estados mentales como la creencia y el deseo cuyos contenidos se expresan convencionalmente mediante cláusulas que utilizan "que"—, el supuesto típico ha sido que las estructuras simbólicas pertinentes son oraciones (Fodor 1975, 1978). La aparente necesidad de reconocer oraciones internas en una explicación de las actitudes proposicionales (en adelante, AP) deriva de varias fuentes distintas.

Para comenzar, las AP son sistemáticas: existen conexiones intrínsecas entre ciertos pensamientos. Consideremos, por ejemplo, el pensamiento de que el niño persigue al perro y el pensamiento de que el perro persigue al niño. Cualquier persona que tenga la capacidad de pensar el primero, tendrá también la capacidad de pensar el segundo, y viceversa. En segundo lugar, las AP son productivas: tenemos la capacidad, al parecer, de pensar indefinidamente una gran cantidad de pensamientos y de formarnos pensamientos que nunca habíamos pensado antes. Las AP son también de alta definición: puedo pensar que tú sabes algo importante sin pensar que tú crees, justificable y verdaderamente, algo importante, aun cuando el análisis correcto de "saber" sea "creer justificable y verdaderamente". Por

[9] La relación de constitución es más débil que la relación de identidad. *A* puede estar constituido por *B* aun cuando *A* y *B* difieran en algunas de sus propiedades modales (Tye 1992).

último, las AP poseen valores de verdad: mi creencia de que la libra esterlina vale menos que hace dos años en relación con el dólar es verdadera o falsa.

Se producen hechos exactamente paralelos a éstos en el caso de oraciones expresadas en lenguaje público. De acuerdo con los partidarios de la hipótesis del lenguaje del pensamiento, estos paralelismos se explican mejor suponiendo que las AP mismas tienen una estructura semejante a una oración.[10]

¿Qué clase de representaciones son entonces los dolores? ¿Son también oraciones internas? Puede resultar tentador suponer que si poseen contenido representacional, como he insistido, y si la hipótesis del lenguaje del pensamiento es verdadera, entonces *deben* ser oraciones. Pero sería una suposición demasiado apresurada. Pues un compromiso con un lenguaje interno dentro del cual se produce la cognición no es un compromiso con oraciones para todos y cada uno de los casos de representaciones mentales. Después de todo, las computadoras —manipuladoras de símbolos *par excellence*— operan sobre todo tipo de estructuras simbólicas (por ejemplo, listas, oraciones, ordenamientos), y, por mencionar sólo un ejemplo, hay conocidas teorías de las imágenes mentales que caen dentro del enfoque computacional,[11] pero que rechazan la tesis de que las imágenes son oraciones. Más aún, indudablemente no resulta obvio que la hipótesis del lenguaje del pensamiento, en su forma irrestricta, sea verdadera. Quizá algunas representaciones mentales no están en lo absoluto codificadas en nuestra cabeza con símbolos lingüísticos.

Una simple referencia a la hipótesis general del lenguaje del pensamiento no justifica la afirmación de que los dolores son oraciones. ¿Apoyan las consideraciones aducidas antes con relación a las AP, la concepción de los dolores como oraciones? Podría sostenerse que, hasta cierto punto, así es. Existen algunas relaciones sistemáticas entre los dolores: la capacidad de sentir un ardor en la pierna y (simultáneamente o en otro momento) una punzada en el brazo parece relacionada con la capacidad de

[10] Para un desarrollo detallado de las consideraciones expuestas anteriormente sobre la sistematicidad y la productividad, véase Fodor y Pylyshyn 1988.

[11] Una de estas teorías es el pictoricismo de Stephen Kosslyn (1980).

sentir una punzada en la pierna y un ardor en el brazo. Además, si los dolores son representaciones, pueden ser representados erróneamente.[12] Por lo tanto, hay un sentido en el cual algunos dolores pueden caracterizarse como falsos: lo que representan no es, de hecho, el caso.

Debe admitirse, sin embargo, que estas consideraciones no son muy convincentes. Las relaciones sistemáticas entre los dolores son limitadas y pueden explicarse fácilmente, por ejemplo, con un modelo de los dolores como mapas. Y la posibilidad de una representación errónea no *exige* que los dolores sean verdaderos o falsos, no más que el hecho de que los mapas o las pinturas [*pictures*] sean a veces inexactos exige que sean verdaderos o falsos.

¿Hay alguna evidencia en contra del enfoque oracional del dolor? Me parece que sí. Sabemos que en la percepción visual, la imagen en la retina se reconstruye en la corteza visual de suerte que en un sentido muy literal las zonas adyacentes de la corteza representan zonas adyacentes de la imagen retinal. Se produce, pues, una ordenada proyección topográfica de la imagen retinal en el cerebro. Esto ha sido establecido mediante experimentos en los que un electrodo detector se coloca en la corteza visual. El electrodo recoge una mayor actividad neuronal cuando se dirige una luz a un punto particular de la retina. Al mover un poco el electrodo, sólo se obtiene el registro continuo de esa mayor actividad si la luz se dirige a una zona adyacente de la retina.

Esta clase de organización topográfica también se encuentra en la corteza somatosensorial. Hay, por ejemplo, una ordenada representación topográfica de la superficie del cuerpo humano especializada en el tacto. En este caso, las regiones adyacentes de la superficie corporal se proyectan sobre regiones adyacentes de la corteza. Una actividad más intensa en una de las regiones corticales pertinentes representa que la región de la superficie corporal proyectada en ella está siendo tocada. Ciertas partes relativamente pequeñas del cuerpo, por ejemplo, las manos y el rostro, contienen entradas para más neuronas que

[12] No son raros los casos de errores en la representación. El dolor en la parte superior del brazo izquierdo suele ser causado por trastornos cardiacos.

algunas partes relativamente extensas, por ejemplo, el tronco. Ésta es la razón de que, cuando se tocan dos puntos separados de la cara, la distancia más corta entre los puntos en la que ambos pueden sentirse es mucho menor que la distancia más corta entre dos puntos tocados en el tronco.

Hay otras representaciones del cuerpo humano en la corteza somatosensorial que están estructuradas de manera semejante. Se ha establecido que la experiencia del dolor está asociada con la actividad en esta corteza.[13] Ahora bien, el hecho de que la corteza somatosensorial esté organizada topográficamente y de que sea la localización primaria del dolor plantea dudas sobre el enfoque oracional del dolor. Pues las oraciones carecen de la estructura representacional en forma de mapa que se requiere.[14]

Mi sugerencia es, pues, que los dolores mismos son representaciones topográficas o semejantes a mapas. En términos más específicos, sostengo que los dolores son patrones de células activas que tienen lugar en ordenamientos tridimensionales, topográficamente estructurados, a los que se anexan etiquetas descriptivas. Esta propuesta puede desarrollarse como sigue.

Para cada dolor hay un ordenamiento conformado por células que se corresponde con partes del cuerpo de tamaño irregular, donde las células adyacentes representan regiones corporales adyacentes.[15] Cada célula, cuando está activa, puede ser concebida como la representación de que algo doloroso está ocurriendo en la región del cuerpo correspondiente. La irregularidad de la definición en el ordenamiento es parcialmente responsable, sugiero, de las variaciones en nuestra experiencia del dolor cuando el mismo grado de daño se produce en regiones corporales diferentes del mismo tamaño (por ejemplo, la cara en contraposición con el torso).

[13] Esto no quiere decir que esté negando que otras regiones neuronales también desempeñen cierto papel en algunas experiencias del dolor. En particular, hay conductos del dolor que terminan tanto en la corteza parietal posterior como en la corteza frontal superior.

[14] Para una discusión de las diferencias representacionales entre las oraciones, las pinturas [*pictures*] y los mapas, véase Tye 1991.

[15] Las características de los ordenamientos se examinan más detenidamente en Tye 1991.

Como las células dentro del propio ordenamiento del dolor se ocupan individualmente sólo de regiones corporales arbitrarias, no se produce ninguna representación en el ordenamiento de las partes naturales del cuerpo. La segmentación de las regiones corporales en tales partes tiene lugar a través de procesos de inspección que examinan los patrones de las células activas en el ordenamiento y que las evalúan, con base en su ubicación, como dolor en brazos, piernas y demás. Es aquí donde las etiquetas descriptivas que representan las partes corporales pertinentes se anexan al ordenamiento. Mi sugerencia es que se introducen otras etiquetas para los rasgos generales del daño representado a través de rutinas ulteriores que funcionan mecánicamente sobre el ordenamiento y que extraen por lo menos *parte* de la información pertinente de sus contenidos. Por ejemplo, en el caso de un dolor punzante, podemos suponer que se produce un repentino patrón de actividad en el ordenamiento, el cual comienza en una zona de este último que representa una limitada región de la superficie corporal y se extiende en la secuencia temporal adecuada hacia células que representan regiones internas adyacentes más profundas (de suerte que se delimita un volumen que tiene la forma aproximada de un puñal). Las rutinas computacionales pertinentes procesan esta actividad y le asignan un término descriptivo apropiado.[16] En el caso de un aguijonazo, podemos suponer que ciertas células del ordenamiento, las cuales representan regiones contiguas de la superficie corporal a lo largo de una angosta franja entran en intensa actividad, más o menos simultáneamente, durante un breve periodo. Esta actividad genera una respuesta computacional, y nuevamente se le atribuye el término pertinente. La posibilidad de que estas suposiciones se encuentren dentro de

[16] No es preciso suponer que la secuencia temporal adecuada a la que se hace referencia en el último enunciado del texto corresponde *necesariamente* a la secuencia temporal del mundo real. El hecho de que las rutinas de la inspección traten la actividad de una célula *C* como si representara un trastorno corporal posterior a la actividad de otra célula *C'* no requiere que se lo instrumente a través de un ordenamiento en el cual *C* se active después de *C'*. Para algunos esclarecedores comentarios sobre la representación del tiempo que pueden servir para apoyar este argumento, véase Dennett 1991, capítulo 6.

los lineamientos correctos es cuestión de investigación de la psicología cognoscitiva.

Desde luego, este burdo modelo es realmente muy esquemático. Sugiero que lo que nos ofrece este modelo es una forma alternativa de pensar los dolores como representaciones, que es más prometedora que el enfoque meramente oracional. Creo que los dolores se representan un poco al estilo en que se representan los mapas que contienen información descriptiva adicional para puntos sobresalientes ("tesoro sepultado aquí", "montaña más elevada de la isla"). A este respecto, se parecen mucho a las imágenes mentales, según yo las concibo.

Hay sólida evidencia de que las imágenes y las percepciones visuales comparten un medio que ha sido denominado "memoria de retención visual". Este medio es de carácter funcional: se compone de un gran número de células, cada una de las cuales se especializa en representar, cuando están cargadas, una diminuta zona de la superficie en una ubicación determinada del campo visual. Para las percepciones visuales y las imágenes residuales, la memoria de retención visual está cargada, en casos normales, de procesos que operan sobre la información contenida en la luz que llega a la vista. Para las imágenes mentales (exceptuando las imágenes residuales), la memoria de retención visual está cargada de procesos generacionales que actúan sobre la información almacenada en la memoria a largo plazo relativa al aspecto de los objetos y a su estructura espacial.

Imágenes y percepciones, como lo he argumentado en otro lugar, son patrones de células cargadas de símbolos que se interpretan en la memoria de retención visual. Los símbolos que están dentro de cada célula representan por lo menos parte de las siguientes características locales: la presencia de una zona superficial, la orientación de la superficie, el color, la textura y así sucesivamente. Las interpretaciones se anexan a los patrones de las células cargados bajo la forma de descripciones que proporcionan un contenido más específico, por ejemplo, si el objeto de la imagen es un círculo o un cuadrado o, en casos más complejos, un pato o un conejo. He formulado más detenidamente esta concepción en otro escrito (Tye 1991), por lo que no voy a desarrollarla aquí. Mi conjetura es que las sensaciones

corporales en general, las experiencias perceptuales y las experiencias con imágenes tienen sus contenidos codificados en ordenamientos del tipo que acabo de describir.

Quizá deba subrayar aquí una idea que presenté antes con relación a las imágenes residuales. Desde mi punto de vista, aunque los procesos responsables de cargar los ordenamientos tanto de la sensación corporal como de la perceptual no requieren esencialmente la creencia o el pensamiento, implican sin duda la categorización. Consideremos, por ejemplo, las clases de categorización que se producen en las primeras etapas de la visión, por ejemplo, en la detección de bordes y el cálculo de la distancia —las categorizaciones que son pertinentes en la manera en que se carga la memoria de retención visual. Se trata de categorizaciones automáticas. No exigen que la criatura tenga creencias o pensamientos sobre las propiedades de los estímulos visuales que están representados en esas categorizaciones. Sostengo que lo mismo es cierto, en gran parte, para el caso de las sensaciones corporales. De este modo, no hay nada en la explicación propuesta que implique que un niño muy pequeño o un animal no puedan sentir dolor. Su relativa pobreza conceptual no les impide experimentar la clase de procesamiento y representaciones que se necesitan para cargar una parte del ordenamiento apropiado.

¿Qué ocurre con las etiquetas descriptivas que se anexan a los ordenamientos? ¿Está involucrado en este caso el pensamiento o la creencia? También aquí me inclino a suponer que, por lo menos en *algunos* casos, las categorizaciones ulteriores que tienen lugar en estas circunstancias se producen sin la participación esencial de conceptos (entendidos como los que entrañan un lenguaje o pensamiento públicos). Consideremos, por ejemplo, el caso de formas simples en la experiencia perceptual. Nada me parecerá cuadrado a menos que los procesos apropiados hayan operado en las células cargadas de la memoria de retención visual, categorizándolas como la representación de una forma cuadrada.[17] Pero no es necesario que yo piense (o crea) que el objeto que estoy viendo es cuadrado. De hecho, no necesito te-

[17] Para un posible bosquejo de estos procesos, véase Marr 1982.

ner absolutamente ningún pensamiento (o creencia) acerca de las características reales o aparentes del objeto observado.

Tal vez se dirá ahora que no queda claro el modo en que el modelo que he reseñado se adecúa al hecho perfectamente establecido de que el dolor está expuesto a influencias de un nivel superior. Por ejemplo, en cierto experimento se vio que los corredores eran más veloces en una encantadora zona boscosa que en una pista. Aparentemente, experimentaban menos dolor en brazos y piernas mientras contemplaban los árboles y las flores y, en consecuencia, corrían a un paso más rápido (Pennebaker y Lightner 1980). También está el interesante caso de unos *terrier* escoceses criados en ambientes restringidos. Una vez liberados, nos cuenta Melzack, se comportaron como sigue:

> Estaban tan retozones y revoltosos que no faltó quien les pisara accidentalmente la cola. Pero no les oímos ni un chillido. En su excitación, también se golpeaban la cabeza dando un sonoro chasquido en la tubería del agua del edificio, que tenía poca altura, y se limitaban a alejarse. Por curiosidad, encendí un cerillo para ver cómo reaccionarían ante la llama. Acercaron el hocico a ella una y otra vez y, aunque retrocedían como ante un olor extraño, no reaccionaron como si les doliera. (Citado en Warga 1987)

La ansiedad, en cambio, intensifica la experiencia del dolor, como, por ejemplo, cuando se compara una lesión presente con otra pasada.

Estos hechos acerca del dolor no implican ninguna amenaza para la propuesta que he formulado. Pueden explicarse mediante el supuesto de que la vía receptora del dolor en la columna vertebral, la cual conduce a la corteza somatosensorial, tiene una entrada controlada por las entradas de los centros cerebrales superiores (la teoría del control de entrada) (Melzack y Wall 1965). Cuando esta entrada se encuentra parcialmente cerrada, pasa menos información y la sensación de dolor disminuye. Cuando se abre más, se permite el paso de una mayor cantidad de información. La ansiedad, la excitación, la alegría, la concentración y otras actividades de nivel superior afectan la orientación de la entrada. De este modo, el hecho de que la experiencia del dolor puede penetrar cognoscitiva-

mente, *en el sentido que se acaba de señalar*, no presenta ninguna dificultad real para mi propuesta.

Aun así, el dolor posee una característica muy importante sobre la que, hasta ahora, sólo he hecho un par de comentarios rápidos: su carácter fenoménico. ¿En qué consiste el carácter fenoménico del dolor? ¿Cuál es la relación entre el carácter fenoménico y el contenido representacional?

Éstas son cuestiones que abordaré en la sección final. Hasta aquí mi interés se ha limitado a establecer la tesis de que los dolores poseen contenido representacional y a exponer de manera admisible mi afirmación de que los dolores, como las imágenes, constituyen un objeto de estudio adecuado para la psicología cognoscitiva. El que los dolores posean características fenoménicas que no puedan fundarse representacional o intencionalmente es un problema distinto (tal como ocurre en el caso de las imágenes mentales).

IV

Se sostiene, por lo general, que los dolores poseen propiedades intrínsecas, introspectivamente accesibles, que son enteramente no intencionales y que son las *únicas* responsables de su carácter fenoménico. Tales propiedades suelen llamarse *qualia*, aunque el término se emplea a veces en un sentido más amplio para referirse a los rasgos subjetivos o fenoménicos, independientemente de la manera en que se los analice o entienda. Como concepción general del carácter fenoménico del dolor, este punto de vista me parece a todas luces erróneo.

Consideremos el caso siguiente: mientras escribo, me duele la espalda. Hay algo que es para mí cómo es ser el sujeto de este dolor de espalda. Cómo es ser el sujeto de este dolor de espalda equivale, desde luego, a cómo es ser el sujeto de este dolor de espalda. Pero cómo es ser el sujeto de este dolor de espalda no equivale a cómo es ser yo, aun cuando yo sea el sujeto de este dolor de espalda. ¿Por qué no?

Sugiero que la respuesta consiste en que el contexto de "cómo es" es intencional. No siempre se pueden sustituir las expresiones correferenciales sin incurrir en el riesgo de cambiar el valor

de verdad. Lo que origina el contexto intensional es la naturaleza intencional del carácter fenoménico o de "sentir". El carácter fenoménico específico de un estado –cómo es experimentarlo– no es otro que el contenido intencional del estado o, más precisamente desde mi punto de vista, un aspecto de ese contenido.

Volvamos a considerar el caso citado del dolor de espalda. Las cualidades que experimento (y que me desagradan profundamente) son experimentadas como rasgos instanciados en cierta región de mi espalda y no como rasgos intrínsecos de mi experiencia. Puesto que podría darse el caso de que realmente no haya ningún problema en mi espalda, las cualidades no necesitan ser rasgos efectivos de mi espalda. Más bien, se trata de rasgos que mi experiencia *representa* como instanciados en mi espalda (por ejemplo, el rasgo de ser tal y cual clase de trastorno). Aún más, estos rasgos no son propiedades intrínsecas de mi experiencia que yo proyecto equivocadamente en una zona de mi cuerpo. No hay ningún error general incluido en las experiencias del dolor.

Ahora bien, el carácter fenoménico de mi dolor es intuitivamente algo que me es dado por vía de la introspección de *qué* experimento cuando tengo ese dolor. Pero lo que experimento es lo que representa mi experiencia. Así, el carácter fenoménico es representacional.

Se negará quizá que las cualidades que experimento cuando tengo un dolor de espalda se encuentren aparentemente localizadas en mi espalda. Pero esto es sin duda lo que sugiere firmemente la introspección. Cuando vuelvo la mirada hacia el interior y trato de centrar mi atención en los rasgos *intrínsecos* de la experiencia de mi dolor en este caso, rasgos que la experiencia posee en sí misma aparte de lo que ella representa, parece que no me topo con ninguno. Más bien, parece que siempre termino prestando atención a lo que estoy experimentando *en la espalda*. Una introspección cuidadosa sólo revela aspectos ulteriores de *eso*, aspectos ulteriores del contenido representacional o intencional de mi experiencia. La experiencia misma es transparente.[18] ¿Por qué? Sugiero que la respuesta consiste en

[18] La transparencia ha sido discutida por varios filósofos. Véase, por ejemplo, Harman 1990.

que mi experiencia carece de rasgos *susceptibles de introspección* que la distingan de otras experiencias aparte de las implicadas en su contenido. Por lo tanto, el carácter fenoménico específico de mi experiencia –que en sí mismo es algo a lo que puede accederse a través de la introspección– es idéntico a su contenido intencional general, o está contenido en él.

Aun así, si el carácter fenoménico distintivo de los dolores constituye un aspecto de su contenido intencional, ¿qué aspecto es éste exactamente? ¿Qué es el *contenido fenoménico*? Se trata de una pregunta compleja que abordo detenidamente en mi libro *Ten Problems of Consciousness* (1995). En este capítulo, sólo puedo hacer algunas observaciones muy esquemáticas.

En mi opinión, el contenido fenoménico es un contenido abstracto, no conceptual e intencional en posición [*poised*] de ser usado por los centros cognoscitivos. La afirmación de que los contenidos pertinentes al carácter fenoménico de los dolores (y otras experiencias sensoriales) tienen que estar *posicionados* [*poised*] ha de entenderse como la exigencia de que tales contenidos estén ligados a las representaciones de salida del o de los módulos sensoriales pertinentes y se encuentren listos y en posición de ejercer un impacto directo en el sistema de creencias y deseos.[19]

Esta concepción implica que ninguna creencia podría tener carácter fenoménico. Un contenido se clasifica como fenoménico sólo si es no conceptual y está posicionado. Las creencias no son no conceptuales, ni se hallan adecuadamente posicionadas. Se encuentran en el ámbito conceptual, en vez de proveer entradas a éste. Las creencias no son en modo alguno representaciones sensoriales (aunque en determinadas ocasiones puedan sin duda hallarse acompañadas por tales representaciones).

La afirmación de que los contenidos pertinentes al carácter fenoménico deben ser *abstractos* ha de entenderse como la exigencia de que no entre ningún objeto concreto en tales contenidos. La razón de este requisito salta a la vista. Ya sea que uno

[19] Esto se desarrolla con mayor profundidad en Tye 1995 y 1996. Para una discusión sobre ciertas diferencias entre la concepción de Block (1995) acerca de lo que significa que un estado esté dispuesto y la mía, véase Tye 1996.

tenga o no pierna izquierda, puede sentir dolor en la pierna izquierda, y en ambos casos el carácter fenoménico de la experiencia puede ser exactamente el mismo. Por lo tanto, no se requiere la existencia de esa pierna particular para ese carácter fenoménico dado. Lo importante es, más bien, la conjunción de rasgos o propiedades generales que representa la experiencia. Ésta representa en forma no conceptual que existen tales y cuales rasgos coinstanciados, de localización y otros, por los que adquiere su carácter fenoménico.

La afirmación de que los contenidos pertinentes al carácter fenoménico deben ser *no conceptuales* ha de entenderse como la afirmación de que los rasgos generales que entran en estos contenidos no requieren que los sujetos posean conceptos correspondientes para ellos. Ya he expresado algunas observaciones relativas a este requisito.

¿Cuáles son exactamente, entre los rasgos representados por los dolores, los que forman parte de sus contenidos fenoménicos? No hay una respuesta *a priori*. Se necesita la investigación empírica. Los rasgos pertinentes serán los representados por las representaciones de salida del módulo sensorial del dolor. Podríamos llamar *rasgos observacionales* a los que están así representados en relación con las salidas de los diversos módulos sensoriales. Puesto que los receptores asociados con los módulos y el procesamiento que se produce en ellos varían, los rasgos que son observacionales para un módulo no tienen necesariamente que ser observacionales para otro.[20] Lo que sale depende evidentemente de lo que entra, y de la manera en que opera el módulo. En mi opinión, es la representación de cierta clase de rasgos observacionales formados por nuestras experiencias del dolor y el papel que éstos desempeñan lo que da a los dolores su carácter fenoménico.

Desde luego, todavía quedan muchas preguntas, de las cuales no son de menor importancia las referentes a la superveniencia

[20] Además, al clasificar un rasgo como observacional, no estoy suponiendo que posee esa condición para todas las especies de criaturas posibles. En mi opinión, el carácter observacional es relativo al equipo sensorial de las criaturas. De este modo, algunos rasgos que son observacionales para nosotros, pueden no serlo para otras criaturas posibles (y viceversa).

del carácter fenoménico sobre lo que hay en la cabeza y a la cuestión de los *qualia* ausentes (Tye 1995). Pero espero haber dicho lo suficiente para mostrar que resulta prometedor abordar los dolores desde un punto de vista enteramente representacional.

[*Traducción de Ana Isabel Stellino*]

BIBLIOGRAFÍA

Armstrong, D., 1962, *Bodily Sensations*, Routledge and Kegan Paul, Londres.

Aune, B., 1967, *Knowledge, Mind, and Nature*, Random House, Nueva York.

Block, N., 1995, "On a Confusion about the Function of Consciousness", *Behavioural and Brain Sciences*, vol. 18, no. 2, pp. 227–247. Reproducido en Block, Flanagan y Guzeldere 1997, pp. 375–415.

——, 1983, "Mental Pictures and Cognitive Science", *Philosophical Review*, vol. 92, pp. 499–542.

Block, N., O. Flanagan, y G. Guzeldere (comps.), 1997, *The Nature of Consciousness*, MIT Press, Cambridge, Mass.

Dennett, D., 1991, *Consciousness Explained*, Little Brown, Boston. [Versión en castellano: *La conciencia explicada*, trad. S. Balari Ravera, Paidós, Barcelona, 1995.]

Fodor, J., 1978, "Propositional Attitudes", *Monist*, vol. 61, pp. 501–523.

——, 1975, *The Language of Thought*, Thomas Crowell, Nueva York. [Versión en castellano: *El lenguaje del pensamiento*, trad. J. Fernández Zuliaca, Alianza, Madrid, 1984.]

——y Z. Pylyshyn, 1988, "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition*, vol. 28, pp. 3–71.

Harman, G., 1990, "The Intrinsic Quality of Experience", en J. Tomberlin (comp.), *Philosophical Perspectives*, vol. 4, Ridgeview, Northridge, California, pp. 31–52.

Kosslyn, S., 1980, *Image and Mind*, Harvard University Press, Cambridge, Mass.

Marr, D., 1982, *Vision*, W.H. Freeman, San Francisco.

Melzack, R., 1973, "How Acupuncture Can Block Pain", *Impact of Science on Society*, vol. 23, no. 1, pp. 1–8.

——, 1961, "The Perception of Pain", *Scientific American*, vol. 204, no. 2, pp. 41–49.

Melzack, R. y P. Wall, 1965, "Pain Mechanisms: A New Theory", *Science*, vol. 150, no. 699, pp. 971–979.

Pennebaker, J. y J. Lightner, 1980, "Competition of Internal and External Information in an Exercise Setting", *Journal of Personality and Social Psychology*, vol. 39, pp. 165–174.

Smart, J., 1959, "Sensations and Brain Processes", *Philosophical Review*, vol. 68, pp. 141–156.

Stephens, L. y G. Graham, 1987, "Minding-Your P's and Q's: Pain and Sensible Qualities", *Noûs*, vol. 21, pp. 395–406.

Tye, M., 1996, "The Function of Consciousness", *Noûs*, vol. 30, no. 3, pp. 287–305.

———, 1995, *Ten Problems of Consciousness*, MIT Press, Cambridge, Mass.

———, 1992, "Naturalism and the Mental", *Mind*, vol. 101, no. 403, pp. 421–441.

———, 1991, *The Imagery Debate*, MIT Press, Cambridge, Mass.

Warga, C., 1987, "Pain's Gatekeeper", *Psychology Today*, agosto, pp. 51–56.

PROCEDENCIA DE LOS TEXTOS

Thomas Nagel, "What Is It Like to Be a Bat?", *Philosophical Review*, vol. 83, 1974, pp. 435-450. © 1974 Cornell University. Traducido y reproducido con autorización del editor y del autor.

Colin McGinn, "Can We Solve the Mind-Body Problem?", *Mind*, vol. XCVIII, no. 391, julio de 1989, pp. 349-346. Traducido y reproducido con autorización del editor y del autor.

Frank Jackson, "Epiphenomenal Qualia", *Philosophical Quarterly*, vol. 32, no. 127, 1982, pp. 127-136. © The Editors of *The Philosophical Quarterly*. Traducido y reproducido con autorización del editor y del autor.

Frank Jackson, "What Mary Didn't Know", *The Journal of Philosophy*, vol. 83, no. 5, mayo de 1986, pp. 291-295. © *The Journal of Philosophy*. Traducido y reproducido con autorización del editor y del autor.

David Lewis, "What Experience Teaches", *Proceedings of the Russellian Society*, vol. 13, University of Sydney, 1988, pp. 29-57. Reproducido con autorización de Stephanie Lewis.

Ned Block y Jerry Fodor, "What Psychological States Are Not", *Philosophical Review*, 1972, vol. 81, pp. 159-181. Reproducido con autorización del editor.

Sydney Shoemaker, "Functionalism and Qualia", *Philosophical Studies*, vol. 27, 1975, pp. 291-315. Reproducido con autorización del editor y del autor. La traducción que aparece en el presente volumen es de la versión publicada en *Identity, Cause and Mind: Philosophical Essays*, Cambridge University Press, Cambridge, pp. 184-205.

Daniel C. Dennett, "Quining Qualia", en A. Marcel y E. Bisiach (comps.), *Consciousness in Contemporary Science*, Oxford University Press, Oxford, 1988, pp. 42–77. Reproducido con autorización de Oxford University Press. La traducción que aparece en el presente volumen es de la versión publicada en Ned Block *et al.*, *The Nature of Consciousness*, MIT Press, Cambridge, Mass., 1997, pp. 619–642.

Gilbert Harman, "The Intrinsic Quality of Experience", *Philosophical Perspectives, 4, Action Theory and Philosophy of Mind*, 1990, compilado por James E. Tomberlin (© Ridgeview Publishing Co., Atascadero, CA), pp. 31–52. Reproducido con autorización de Ridgeview Publishing Company, y del autor.

Ned Block, "Inverted Earth", *Philosophical Perspectives, 4, Action Theory and Philosophy of Mind*, 1990, compilado por James E. Tomberlin (© Ridgeview Publishing Co., Atascadero, CA), pp. 53–79. Reproducido con autorización de Ridgeview Publishing Company.

Michael Tye, "A Representational Theory of Pains and Their Phenomenal Character", *Philosophical Perspectives, 9, AI, Connectionism, and Philosophical Psychology*, 1995, compilado por James E. Tomberlin (© Ridgeview Publishing Co., Atascadero, CA), pp. 223–240. Reproducido con autorización de Ridgeview Publishing Company, y del autor. La traducción que aparece en el presente volumen es de la versión publicada en Ned Block *et al.*, *The Nature of Consciousness*, MIT Press, Cambridge, Mass., 1997, pp. 329–340.

ÍNDICE ANALÍTICO

carácter intrínseco de la experiencia, 31, 130, 275; carácter cualitativo, 26-27, 31, 39, 184, 187-194, 196, 204, 207-209, 215; cualidad intrínseca de la experiencia, 31, 263-287
cerebro, 5, 16-17, 36, 45, 49, 53, 55, 57, 60, 65-67, 71-72, 74-75, 77-82, 84-87, 89-90, 96-99, 104-105, 107, 120, 138, 220, 224, 255-256, 264, 269n, 273, 275, 306, 308-309, 312, 314-315, 317-318, 339
cómo es, 10, 14, 16, 20, 32, 34, 50, 53, 60, 112, 119-121, 123, 125-130, 132, 135, 137, 141-143, 146-152, 265, 278-280, 285, 290, 293n, 298n, 320, 346; cómo es ser, 12-13, 19, 34, 45-63, 102-104, 120, 129, 137, 141, 345; cómo se experimenta, 7, 10, 12, 57; cómo se siente, 11-12, 14, 16, 19-20, 24, 30, 186
concepto(s), 9, 10-11, 13, 16-17, 28-29, 32, 39, 45-47, 50-52, 54-56, 60-61, 67, 69, 73-77, 80-86, 89,91, 161, 215-216, 232, 236, 241, 244, 247-248, 278-280, 293, 296-299, 309, 319-321, 330, 334, 336, 343, 348; de dolor, 10, 11, 13, 37, 335; individuación de, 9n; teórico(s), 16-17, 29, 32, 77, 80, 86
conciencia, 7, 12n, 16-17, 28-29, 31, 37, 45-47, 65-67, 71-72, 74-82, 84-90, 92, 99, 101, 104, 189-190, 201-203, 207, 210, 214, 220, 241n, 245-246, 264, 267, 275, 284-285, 319, 334; fenoménica, 7-8, 9n, 17, 21, 27-28, 30n, 36, 39
conocimiento, 5, 13, 15, 18, 19, 20, 25, 32-33, 39, 53, 82, 86-87, 100, 102-103, 108, 111-117, 126-128, 143, 148-149, 151, 190, 192-193, 198, 227, 231-232, 249, 320-321, 325; argumento del, 17-21, 32, 39, 96, 97-100, 102-103, 112-114 116-117, 132-136, 140; intensionalidad del, 19, 112
contenido intencional/representacional, 31, 34, 36, 39-40, 266, 276, 284, 290-292, 294, 297-298, 304, 306, 311-315,

320, 325-326, 336, 338, 345-347
creencia(s), 5-6, 7n, 12n, 22, 24, 26, 30, 33, 36-39, 47-48, 51, 88, 97, 151, 156, 159, 165, 174, 176, 190, 201-204, 206-208, 229, 264, 281-283, 285-286, 298, 306, 321, 326, 330, 334, 337-338, 343-344, 347; cualitativa(s), 24-25, 27, 30, 190, 193-194, 201-203, 206-209

datos sensoriales, 218, 267, 272-273, 284
designadores rígidos, 15n, 23, 24, 192n, 197-198
dolor(es), 5, 7-8, 10-12, 14, 16, 23-25, 31, 35-40, 50, 57n, 74n, 90, 96, 104, 158-159, 169-172, 185-194, 207-208, 210, 265, 273-274, 294n, 326-350

ecolocalización, 12, 49; experiencia(s) ecolocalizadora(s), 13, 28n
eliminativismo, 27-31, 86, 87n, 223n
espectro invertido, 21-27, 29, 33-34, 90, 169, 185, 194, 196-202, 206, 209, 223-224, 226, 280-285, 289-292, 294, 295-299, 300-318, 319
experiencia, 5-7, 13-15, 18-19, 22-23, 27, 28n, 29-31, 34-40, 46-51, 54-56, 60-61, 66, 76, 77n, 80, 85n, 86, 96, 98-99, 102-104, 113-117, 119-153, 199-200, 203, 210, 214, 217, 220-221, 223n 226, 229, 231, 236-239, 241-242, 247-248, 250-254, 257-258, 265, 270, 272-274, 277-278, 281, 284-285, 290-293, 296, 298-299, 301-302, 304-307, 309-315, 317-318, 320, 327, 333-335, 340, 343-344, 346-347; consciente, 12, 46, 201, 215, 219, 243, 263, 276; objeto intencional de la, 266-270; visual, 26, 75, 100, 194, 213, 224, 268, 271, 279, 319
explicación(es), 5-6, 7n, 8-11, 14, 17, 38-39, 46-47, 57n, 61, 70-72, 79-83, 85, 86n, 70-72, 79-83, 85, 86n, 128, 142, 151, 164, 167, 170, 191, 194-195, 203, 207, 208n, 209-210, 221n, 225, 242, 249, 252, 275-276, 278, 299, 325, 329-330, 332, 335-337, 343; causales, 8, 13-15; constitutivas o composicionales, 8, 14, 15, 23, 35; hueco explicativo, 14-16, 27, 29, 38-40

fenomenología, 7, 10, 13-15n, 16, 22-26, 39, 51, 54, 60-61, 65, 73n, 74, 77n, 335; características fenomenológicas, 11, 47
fisicalismo, 11, 14-15, 17-19, 21, 29, 35, 47, 57-59, 95-96, 99-101, 102n, 103, 108, 113-114, 116-118, 155-158, 163-164, 171, 178-179, 208
físico; información física, 18-19, 128-130, 132-134, 140-141; fenómenos físicos, 5, 17, 54, 61n, 143; tipos físicos, 14; propiedades físicas, 14, 15, 312
funcionalismo, 15n, 21-27, 29, 31-35, 66, 171, 179, 183-212, 220, 231, 263-265, 274-277, 280, 283, 285-286, 289-290, 299n, 300n, 310, 312, 314-316,

320; funcionalista(s), 22–24, 74n, 185, 187–188, 204, 206, 290, 297, 313, 315–316

habilidades/capacidades, 5, 12n, 13, 20, 32, 115–116, 145–152, 219n, 225, 233–234, 245–247, 255, 257–258, 280, 307, 320, 325, 330, 337–338; cognoscitivas 15, 325; recognoscitivas 20
hipótesis de los *qualia* ausentes, 21–22

identidad, 11, 14, 21, 23–24, 26–27, 124, 127–128, 156, 158, 161–164, 167, 169–173, 177–180, 183–184, 185n, 186n, 187–189, 193, 198, 204, 207–210, 247, 290n, 301n, 307–308, 311–312, 316, 326–327, 337n
imagen residual, 274, 326–328, 343
inefabilidad, 28n, 250, 254, 257
introspección, 16–17, 25, 28n, 29, 190, 192, 200, 335, 346–347

mente; estar cerrado cognoscitivamente, 16–17, 67, 71, 83–84

objetivo, 9, 15, 30, 38–39, 48, 52–56, 58, 60–61, 70–71, 77n, 86–87, 215, 220, 232, 238, 250n

percepción, 6–7, 16–17, 20, 32–33, 35–36, 195n, 202, 235, 242–244, 246–247, 267–268, 273, 275, 277–280, 282–283, 285, 320, 325, 339, 342; aparato sonar, 12; estado(s) perceptual(es), 5, 26, 281, 310; experiencia perceptual, 6, 12, 18, 30, 96, 210, 251, 266, 270, 276, 343; sistema/aparato perceptual, 10, 12, 14
perspectiva/punto de vista, 9–14, 19, 25, 27, 29, 38–39, 48, 52–56, 60, 96, 104, 128, 146, 273, 278, 285, 294, 299, 309
principio de homogeneidad, 17, 80
propuesta representacionista, 35–40

qualia/quale, 7, 9n, 21–31, 33–36, 39, 96–97, 114–117, 131, 170, 183–212, 217–223, 227–244, 245–248, 250, 253, 255–256, 289–291, 295–298, 302–303, 312, 318, 320–321; ausentes, 21–27, 34, 131–132, 171, 184, 186–188, 190, 192, 201, 207–208, 290n, 297n, 316n, 349; directamente accesibles, 28–29, 220–221, 226, 253; falacia de la intencionalización de, 33, 293–294, 301n, 313, 315n; inefable(s), 28, 219–221, 248, 251, 258. *Véase también* inefabilidad; intrínsecos, 28, 215, 221, 251, 254, 258; privados, 28, 221, 239n, 254, 258; zombi(s), 22–23, 25, 290

representación, 20, 31–32, 36n, 37–38, 51, 151, 168, 174, 267, 269, 272, 278–279, 281–284, 309, 327, 334n, 335–336, 338–343, 347–348

sensación, 5–10, 12, 16, 18, 20–21n, 31, 35–40, 86–87, 96, 100, 112, 114, 129–130, 165, 184, 186–187, 190, 192, 210, 215, 241n, 263–265, 291, 296, 307,

317-318, 326, 330, 334, 336, 342-344

similitud y diferencia cualitativas, 26, 187, 198, 201, 203-204, 206, 210, 221, 222n, 309-310

subjetivo(s), 7-9, 11-13, 15, 27, 31, 38-39, 46-49, 51-52, 54n, 55-56, 57n, 59-61, 77n, 146, 218, 222-223, 239-240, 243, 345

superveniencia, 14-15, 23, 89-90, 101n, 133-136, 312, 348

teoría de identidad de los estados funcionales/TIEF, 21-24, 26-27, 161-171, 173, 175n, 176, 179, 183-184, 187, 307

teoría representacional, 36, 325-350

Tierra Invertida, 33-34, 289-323

visión ciega, 6n

ÍNDICE

MAITE EZCURDIA y OLBETH HANSBERG, Sensaciones, conciencia fenoménica y los límites de la explicación.. 5

THOMAS NAGEL, ¿Cómo es ser un murciélago? 45

COLIN MCGINN, ¿Podemos resolver el problema mente-cuerpo? 65

FRANK JACKSON, *Qualia* epifenoménicos 95

FRANK JACKSON, Lo que María no sabía 111

DAVID LEWIS, Lo que enseña la experiencia 119

NED BLOCK y JERRY FODOR, Lo que no son los estados psicológicos 155

SYDNEY SHOEMAKER, Funcionalismo y *qualia* 183

DANIEL DENNETT, Quinear los *qualia* 213

GILBERT HARMAN, La cualidad intrínseca de la experiencia 263

NED BLOCK, La Tierra Invertida 289

MICHAEL TYE, Una teoría representacional del dolor y de su carácter fenoménico 325

Procedencia de los textos 351

Índice analítico 353

La naturaleza de la experiencia se terminó de imprimir el 12 de agosto de 2003 en Formación gráfica S.A. de C.V., Matamoros 112, Col. Raúl Romero, 57630, Cd. Nezahualcoyótl, Edo. de México. En su impresión se utilizó papel cultural de 90 gramos. Para su composición y formación, realizadas por computadora en el Instituto de Investigaciones Filosóficas, se utilizaron el programa TEX y tipos Newbaskerville. El cuidado de la edición estuvo a cargo de Claudia Chávez A.

El tiraje consta de 500 ejemplares.